

This Page Is Inserted by IFW Operations  
and is not a part of the Official Record

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning documents *will not* correct images,  
please do not report the images to the  
Image Problem Mailbox.**

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 28, 2003, 09:01:30 ; Search time 68.7874 Seconds  
(without alignments)  
251.615 Million cell updates/sec

Sequence: 1 QGRGSSQSIISPMRSEIN.....SPTASSOSATNMAIHRSP 84

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 74486

Minimum DB seq length: 40  
Maximum DB seq length: 90

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 1000 summaries

- 1: sp\_bacteria.\*
- 2: sp\_fungi.\*
- 3: sp\_human.\*
- 4: sp\_invertebrate.\*
- 5: sp\_mammal.\*
- 6: sp\_mhc.\*
- 7: sp\_organelle.\*
- 8: sp\_phase.\*
- 9: sp\_plant.\*
- 10: sp\_rodent.\*
- 11: sp\_virus.\*
- 12: sp\_vertebrate.\*
- 13: sp\_unclassified.\*
- 14: sp\_rvirus.\*
- 15: sp\_bacteriap.\*
- 16: sp\_archaea.\*
- 17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	62.5	14.8	67	16	Q93J23
2	58	13.7	74	4	Q8WUP6
3	56	13.2	58	9	Q80316
4	54	12.8	80	5	Q23341
5	54	12.8	84	16	Q9JMT2
6	53	12.5	80	12	Q91LH8
7	51	12.1	88	13	Q57347
8	48.5	11.5	65	5	Q9NFC8
9	48.5	11.5	90	7	Q9BD73
10	48	11.3	89	16	Q9R6C9
11	47.5	11.2	66	15	Q93D08
12	47.5	11.2	79	16	Q922Q9
13	47	11.1	86	2	Q30681
14	47	11.1	86	2	Q30681
15	47	11.1	86	2	Q30681
16	46.5	11.0	79	2	Q9RCD4

Q96u90 neurospora  
Q98kz6 rhizobium l  
Q93wv8 narcissus p  
Q8vy75 arabidopsis  
Q76048 human immun  
Q8wvf9 sus scrofa  
Q9phk1 xylella fas  
Q9x14 bacterioph  
Q9pcx0 xylella fas  
P91302 caenorhabdi  
P94624 clostridium  
Q89224 vaccinia vi  
Q91521 human immun  
Q8902 archaeoglob  
Q91q4 neisseria m  
Q91u7 pyrus pyrifi  
Q95mb4 pan troglod  
Q95mb3 gorilla gor  
Q95mb2 pongo pygma  
Q95mb1 papio anubi  
Q95mb0 macaca mula  
Q95mb9 macaca sile  
Q27686 methanobact  
Q00448 pisum sativ  
Q8yq5 anabaena sp  
Q91124 human immun  
Q8vn35 helicobacte  
Q8vn29 helicobacte  
P73181 synechocyst  
Q8tvm9 methanopyru  
Q8ttk0 methanosarc  
Q16833 drosophila  
Q91136 human immun  
Q91162 human immun  
Q91156 human immun  
Q91155 human immun  
Q91154 human immun  
Q8x555 escherichia  
Q8x554 escherichia  
Q8x553 escherichia  
Q8x552 escherichia  
Q8x551 escherichia  
Q8x550 escherichia  
Q8x549 escherichia  
Q8x548 escherichia  
Q8x547 escherichia  
Q8x546 escherichia  
Q8x545 escherichia  
Q8x544 escherichia  
Q8x543 escherichia  
Q8x542 escherichia  
Q8x541 escherichia  
Q8x540 escherichia  
Q8x539 escherichia  
Q8x538 escherichia  
Q8x537 escherichia  
Q8x536 escherichia  
Q8x535 escherichia  
Q8x534 escherichia  
Q8x533 escherichia  
Q8x532 escherichia  
Q8x531 escherichia  
Q8x530 escherichia  
Q8x529 escherichia  
Q8x528 escherichia  
Q8x527 escherichia  
Q8x526 escherichia  
Q8x525 escherichia  
Q8x524 escherichia  
Q8x523 escherichia  
Q8x522 escherichia  
Q8x521 escherichia  
Q8x520 escherichia  
Q8x519 escherichia  
Q8x518 escherichia  
Q8x517 escherichia  
Q8x516 escherichia  
Q8x515 escherichia  
Q8x514 escherichia  
Q8x513 escherichia  
Q8x512 escherichia  
Q8x511 escherichia  
Q8x510 escherichia  
Q8x509 escherichia  
Q8x508 escherichia  
Q8x507 escherichia  
Q8x506 escherichia  
Q8x505 escherichia  
Q8x504 escherichia  
Q8x503 escherichia  
Q8x502 escherichia  
Q8x501 escherichia  
Q8x500 escherichia  
Q8x499 escherichia  
Q8x498 escherichia  
Q8x497 escherichia  
Q8x496 escherichia  
Q8x495 escherichia  
Q8x494 escherichia  
Q8x493 escherichia  
Q8x492 escherichia  
Q8x491 escherichia  
Q8x490 escherichia  
Q8x489 escherichia  
Q8x488 escherichia  
Q8x487 escherichia  
Q8x486 escherichia  
Q8x485 escherichia  
Q8x484 escherichia  
Q8x483 escherichia  
Q8x482 escherichia  
Q8x481 escherichia  
Q8x480 escherichia  
Q8x479 escherichia  
Q8x478 escherichia  
Q8x477 escherichia  
Q8x476 escherichia  
Q8x475 escherichia  
Q8x474 escherichia  
Q8x473 escherichia  
Q8x472 escherichia  
Q8x471 escherichia  
Q8x470 escherichia  
Q8x469 escherichia  
Q8x468 escherichia  
Q8x467 escherichia  
Q8x466 escherichia  
Q8x465 escherichia  
Q8x464 escherichia  
Q8x463 escherichia  
Q8x462 escherichia  
Q8x461 escherichia  
Q8x460 escherichia  
Q8x459 escherichia  
Q8x458 escherichia  
Q8x457 escherichia  
Q8x456 escherichia  
Q8x455 escherichia  
Q8x454 escherichia  
Q8x453 escherichia  
Q8x452 escherichia  
Q8x451 escherichia  
Q8x450 escherichia  
Q8x449 escherichia  
Q8x448 escherichia  
Q8x447 escherichia  
Q8x446 escherichia  
Q8x445 escherichia  
Q8x444 escherichia  
Q8x443 escherichia  
Q8x442 escherichia  
Q8x441 escherichia  
Q8x440 escherichia  
Q8x439 escherichia  
Q8x438 escherichia  
Q8x437 escherichia  
Q8x436 escherichia  
Q8x435 escherichia  
Q8x434 escherichia  
Q8x433 escherichia  
Q8x432 escherichia  
Q8x431 escherichia  
Q8x430 escherichia  
Q8x429 escherichia  
Q8x428 escherichia  
Q8x427 escherichia  
Q8x426 escherichia  
Q8x425 escherichia  
Q8x424 escherichia  
Q8x423 escherichia  
Q8x422 escherichia  
Q8x421 escherichia  
Q8x420 escherichia  
Q8x419 escherichia  
Q8x418 escherichia  
Q8x417 escherichia  
Q8x416 escherichia  
Q8x415 escherichia  
Q8x414 escherichia  
Q8x413 escherichia  
Q8x412 escherichia  
Q8x411 escherichia  
Q8x410 escherichia  
Q8x409 escherichia  
Q8x408 escherichia  
Q8x407 escherichia  
Q8x406 escherichia  
Q8x405 escherichia  
Q8x404 escherichia  
Q8x403 escherichia  
Q8x402 escherichia  
Q8x401 escherichia  
Q8x400 escherichia  
Q8x399 escherichia  
Q8x398 escherichia  
Q8x397 escherichia  
Q8x396 escherichia  
Q8x395 escherichia  
Q8x394 escherichia  
Q8x393 escherichia  
Q8x392 escherichia  
Q8x391 escherichia  
Q8x390 escherichia  
Q8x389 escherichia  
Q8x388 escherichia  
Q8x387 escherichia  
Q8x386 escherichia  
Q8x385 escherichia  
Q8x384 escherichia  
Q8x383 escherichia  
Q8x382 escherichia  
Q8x381 escherichia  
Q8x380 escherichia  
Q8x379 escherichia  
Q8x378 escherichia  
Q8x377 escherichia  
Q8x376 escherichia  
Q8x375 escherichia  
Q8x374 escherichia  
Q8x373 escherichia  
Q8x372 escherichia  
Q8x371 escherichia  
Q8x370 escherichia  
Q8x369 escherichia  
Q8x368 escherichia  
Q8x367 escherichia  
Q8x366 escherichia  
Q8x365 escherichia  
Q8x364 escherichia  
Q8x363 escherichia  
Q8x362 escherichia  
Q8x361 escherichia  
Q8x360 escherichia  
Q8x359 escherichia  
Q8x358 escherichia  
Q8x357 escherichia  
Q8x356 escherichia  
Q8x355 escherichia  
Q8x354 escherichia  
Q8x353 escherichia  
Q8x352 escherichia  
Q8x351 escherichia  
Q8x350 escherichia  
Q8x349 escherichia  
Q8x348 escherichia  
Q8x347 escherichia  
Q8x346 escherichia  
Q8x345 escherichia  
Q8x344 escherichia  
Q8x343 escherichia  
Q8x342 escherichia  
Q8x341 escherichia  
Q8x340 escherichia  
Q8x339 escherichia  
Q8x338 escherichia  
Q8x337 escherichia  
Q8x336 escherichia  
Q8x335 escherichia  
Q8x334 escherichia  
Q8x333 escherichia  
Q8x332 escherichia  
Q8x331 escherichia  
Q8x330 escherichia  
Q8x329 escherichia  
Q8x328 escherichia  
Q8x327 escherichia  
Q8x326 escherichia  
Q8x325 escherichia  
Q8x324 escherichia  
Q8x323 escherichia  
Q8x322 escherichia  
Q8x321 escherichia  
Q8x320 escherichia  
Q8x319 escherichia  
Q8x318 escherichia  
Q8x317 escherichia  
Q8x316 escherichia  
Q8x315 escherichia  
Q8x314 escherichia  
Q8x313 escherichia  
Q8x312 escherichia  
Q8x311 escherichia  
Q8x310 escherichia  
Q8x309 escherichia  
Q8x308 escherichia  
Q8x307 escherichia  
Q8x306 escherichia  
Q8x305 escherichia  
Q8x304 escherichia  
Q8x303 escherichia  
Q8x302 escherichia  
Q8x301 escherichia  
Q8x300 escherichia  
Q8x299 escherichia  
Q8x298 escherichia  
Q8x297 escherichia  
Q8x296 escherichia  
Q8x295 escherichia  
Q8x294 escherichia  
Q8x293 escherichia  
Q8x292 escherichia  
Q8x291 escherichia  
Q8x290 escherichia  
Q8x289 escherichia  
Q8x288 escherichia  
Q8x287 escherichia  
Q8x286 escherichia  
Q8x285 escherichia  
Q8x284 escherichia  
Q8x283 escherichia  
Q8x282 escherichia  
Q8x281 escherichia  
Q8x280 escherichia  
Q8x279 escherichia  
Q8x278 escherichia  
Q8x277 escherichia  
Q8x276 escherichia  
Q8x275 escherichia  
Q8x274 escherichia  
Q8x273 escherichia  
Q8x272 escherichia  
Q8x271 escherichia  
Q8x270 escherichia  
Q8x269 escherichia  
Q8x268 escherichia  
Q8x267 escherichia  
Q8x266 escherichia  
Q8x265 escherichia  
Q8x264 escherichia  
Q8x263 escherichia  
Q8x262 escherichia  
Q8x261 escherichia  
Q8x260 escherichia  
Q8x259 escherichia  
Q8x258 escherichia  
Q8x257 escherichia  
Q8x256 escherichia  
Q8x255 escherichia  
Q8x254 escherichia  
Q8x253 escherichia  
Q8x252 escherichia  
Q8x251 escherichia  
Q8x250 escherichia  
Q8x249 escherichia  
Q8x248 escherichia  
Q8x247 escherichia  
Q8x246 escherichia  
Q8x245 escherichia  
Q8x244 escherichia  
Q8x243 escherichia  
Q8x242 escherichia  
Q8x241 escherichia  
Q8x240 escherichia  
Q8x239 escherichia  
Q8x238 escherichia  
Q8x237 escherichia  
Q8x236 escherichia  
Q8x235 escherichia  
Q8x234 escherichia  
Q8x233 escherichia  
Q8x232 escherichia  
Q8x231 escherichia  
Q8x230 escherichia  
Q8x229 escherichia  
Q8x228 escherichia  
Q8x227 escherichia  
Q8x226 escherichia  
Q8x225 escherichia  
Q8x224 escherichia  
Q8x223 escherichia  
Q8x222 escherichia  
Q8x221 escherichia  
Q8x220 escherichia  
Q8x219 escherichia  
Q8x218 escherichia  
Q8x217 escherichia  
Q8x216 escherichia  
Q8x215 escherichia  
Q8x214 escherichia  
Q8x213 escherichia  
Q8x212 escherichia  
Q8x211 escherichia  
Q8x210 escherichia  
Q8x209 escherichia  
Q8x208 escherichia  
Q8x207 escherichia  
Q8x206 escherichia  
Q8x205 escherichia  
Q8x204 escherichia  
Q8x203 escherichia  
Q8x202 escherichia  
Q8x201 escherichia  
Q8x200 escherichia  
Q8x199 escherichia  
Q8x198 escherichia  
Q8x197 escherichia  
Q8x196 escherichia  
Q8x195 escherichia  
Q8x194 escherichia  
Q8x193 escherichia  
Q8x192 escherichia  
Q8x191 escherichia  
Q8x190 escherichia  
Q8x189 escherichia  
Q8x188 escherichia  
Q8x187 escherichia  
Q8x186 escherichia  
Q8x185 escherichia  
Q8x184 escherichia  
Q8x183 escherichia  
Q8x182 escherichia  
Q8x181 escherichia  
Q8x180 escherichia  
Q8x179 escherichia  
Q8x178 escherichia  
Q8x177 escherichia  
Q8x176 escherichia  
Q8x175 escherichia  
Q8x174 escherichia  
Q8x173 escherichia  
Q8x172 escherichia  
Q8x171 escherichia  
Q8x170 escherichia  
Q8x169 escherichia  
Q8x168 escherichia  
Q8x167 escherichia  
Q8x166 escherichia  
Q8x165 escherichia  
Q8x164 escherichia  
Q8x163 escherichia  
Q8x162 escherichia  
Q8x161 escherichia  
Q8x160 escherichia  
Q8x159 escherichia  
Q8x158 escherichia  
Q8x157 escherichia  
Q8x156 escherichia  
Q8x155 escherichia  
Q8x154 escherichia  
Q8x153 escherichia  
Q8x152 escherichia  
Q8x151 escherichia  
Q8x150 escherichia  
Q8x149 escherichia  
Q8x148 escherichia  
Q8x147 escherichia  
Q8x146 escherichia  
Q8x145 escherichia  
Q8x144 escherichia  
Q8x143 escherichia  
Q8x142 escherichia  
Q8x141 escherichia  
Q8x140 escherichia  
Q8x139 escherichia  
Q8x138 escherichia  
Q8x137 escherichia  
Q8x136 escherichia  
Q8x135 escherichia  
Q8x134 escherichia  
Q8x133 escherichia  
Q8x132 escherichia  
Q8x131 escherichia  
Q8x130 escherichia  
Q8x129 escherichia  
Q8x128 escherichia  
Q8x127 escherichia  
Q8x126 escherichia  
Q8x125 escherichia  
Q8x124 escherichia  
Q8x123 escherichia  
Q8x122 escherichia  
Q8x121 escherichia  
Q8x120 escherichia  
Q8x119 escherichia  
Q8x118 escherichia  
Q8x117 escherichia  
Q8x116 escherichia  
Q8x115 escherichia  
Q8x114 escherichia  
Q8x113 escherichia  
Q8x112 escherichia  
Q8x111 escherichia  
Q8x110 escherichia  
Q8x109 escherichia  
Q8x108 escherichia  
Q8x107 escherichia  
Q8x106 escherichia  
Q8x105 escherichia  
Q8x104 escherichia  
Q8x103 escherichia  
Q8x102 escherichia  
Q8x101 escherichia  
Q8x100 escherichia  
Q8x099 escherichia  
Q8x098 escherichia  
Q8x097 escherichia  
Q8x096 escherichia  
Q8x095 escherichia  
Q8x094 escherichia  
Q8x093 escherichia  
Q8x092 escherichia  
Q8x091 escherichia  
Q8x090 escherichia  
Q8x089 escherichia  
Q8x088 escherichia  
Q8x087 escherichia  
Q8x086 escherichia  
Q8x085 escherichia  
Q8x084 escherichia  
Q8x083 escherichia  
Q8x082 escherichia  
Q8x081 escherichia  
Q8x080 escherichia  
Q8x079 escherichia  
Q8x078 escherichia  
Q8x077 escherichia  
Q8x076 escherichia  
Q8x075 escherichia  
Q8x074 escherichia  
Q8x073 escherichia  
Q8x072 escherichia  
Q8x071 escherichia  
Q8x070 escherichia  
Q8x069 escherichia  
Q8x068 escherichia  
Q8x067 escherichia  
Q8x066 escherichia  
Q8x065 escherichia  
Q8x064 escherichia  
Q8x063 escherichia  
Q8x062 escherichia  
Q8x061 escherichia  
Q8x060 escherichia  
Q8x059 escherichia  
Q8x058 escherichia  
Q8x057 escherichia  
Q8x056 escherichia  
Q8x055 escherichia  
Q8x054 escherichia  
Q8x053 escherichia  
Q8x052 escherichia  
Q8x051 escherichia  
Q8x050 escherichia  
Q8x049 escherichia  
Q8x048 escherichia  
Q8x047 escherichia  
Q8x046 escherichia  
Q8x045 escherichia  
Q8x044 escherichia  
Q8x043 escherichia  
Q8x042 escherichia  
Q8x041 escherichia  
Q8x040 escherichia  
Q8x039 escherichia  
Q8x038 escherichia  
Q8x037 escherichia  
Q8x036 escherichia  
Q8x035 escherichia  
Q8x034 escherichia  
Q8x033 escherichia  
Q8x032 escherichia  
Q8x031 escherichia  
Q8x030 escherichia  
Q8x029 escherichia  
Q8x028 escherichia  
Q8x027 escherichia  
Q8x026 escherichia  
Q8x025 escherichia  
Q8x024 escherichia  
Q8x023 escherichia  
Q8x022 escherichia  
Q8x021 escherichia  
Q8x020 escherichia  
Q8x019 escherichia  
Q8x018 escherichia  
Q8x017 escherichia  
Q8x016 escherichia  
Q8x015 escherichia  
Q8x014 escherichia  
Q8x013 escherichia  
Q8x012 escherichia  
Q8x011 escherichia  
Q8x010 escherichia  
Q8x009 escherichia  
Q8x008 escherichia  
Q8x007 escherichia  
Q8x006 escherichia  
Q8x005 escherichia  
Q8x004 escherichia  
Q8x003 escherichia  
Q8x002 escherichia  
Q8x001 escherichia  
Q8x000 escherichia

90	42	9.9	67	15	Q9J194	Q9J194 human immun	163	40.5	9.6	69	2	Q92EU9	Q9zeu9 lactobacill
91	42	9.9	67	15	Q9J172	Q9J172 human immun	164	40.5	9.6	69	15	Q9WMR4	Q9wmr4 human immun
92	42	9.9	67	15	Q9J163	Q9J163 human immun	165	40.5	9.6	69	15	Q9WMO5	Q9wmq5 human immun
93	42	9.9	67	15	Q9J162	Q9J162 human immun	166	40.5	9.6	69	16	Q9KY84	Q9ky84 streptomyce
94	42	9.9	67	15	Q9J160	Q9J160 human immun	167	40.5	9.6	70	7	Q46870	Q46870 homo sapien
95	42	9.9	68	15	Q9J156	Q9J156 human immun	168	40.5	9.6	70	14	Q99IX6	Q99ix6 uncultured
96	42	9.9	68	15	Q9J148	Q9J148 human immun	169	40.5	9.6	70	16	P71998	P71998 mycobacteri
97	42	9.9	68	15	Q9J145	Q9J145 human immun	170	40.5	9.6	74	12	Q66059	Q66059 canine herp
98	42	9.9	68	15	Q9J143	Q9J143 human immun	171	40.5	9.6	76	2	Q8VN33	Q8vn33 helicobacte
99	42	9.9	68	15	Q9J608	Q9J608 human immun	172	40.5	9.6	77	15	Q38442	Q38442 human immun
100	42	9.9	69	16	Q8XNS4	Q8xns4 clostridium	173	40.5	9.6	77	16	Q8YF42	Q8yf42 brucella me
101	42	9.9	72	2	Q48506	Q48506 lactococcus	174	40.5	9.6	78	4	Q96T74	Q96t74 homo sapien
102	42	9.9	72	2	Q07693	Q07693 mycobacteri	175	40.5	9.6	79	2	Q936T5	Q936t5 pseudomonas
103	42	9.9	72	5	Q9VGM8	Q9vgm8 drosophila	176	40.5	9.6	79	17	Q8TUT4	Q8tut4 methanopyru
104	42	9.9	73	17	Q27513	Q27513 methanobact	177	40.5	9.6	81	2	Q9EUC0	Q9euc0 pseudomonas
105	42	9.9	75	4	Q9UGQ6	Q9ugq6 homo sapien	178	40.5	9.6	86	15	Q38069	Q38069 human immun
106	42	9.9	76	12	Q64944	Q64944 avian infec	179	40.5	9.6	86	15	Q38245	Q38245 human immun
107	42	9.9	76	12	Q64947	Q64947 avian infec	180	40.5	9.6	86	15	Q38246	Q38246 human immun
108	42	9.9	79	16	Q928X5	Q928x5 chlamydia p	181	40.5	9.6	86	15	Q38247	Q38247 human immun
109	42	9.9	79	16	Q9JSH8	Q9jsh8 chlamydia p	182	40.5	9.6	86	15	Q38249	Q38249 human immun
110	42	9.9	81	6	Q28836	Q28836 ovis aries	183	40.5	9.6	86	15	Q38250	Q38250 human immun
111	42	9.9	81	16	Q9K247	Q9k247 chlamydia p	184	40.5	9.6	86	15	Q38252	Q38252 human immun
112	42	9.9	85	16	Q9R9D7	Q8r9d7 thermoanaer	185	40.5	9.6	86	15	Q38253	Q38253 human immun
113	42	9.9	86	16	Q92CF9	Q92cf9 listeria in	186	40.5	9.6	86	15	Q38254	Q38254 human immun
114	42	9.9	86	16	Q8Y7M9	Q8y7m9 listeria mo	187	40.5	9.6	86	15	Q38255	Q38255 human immun
115	42	9.9	88	5	Q9S532	Q9s532 drosophila	188	40.5	9.6	86	15	Q38256	Q38256 human immun
116	42	9.9	89	2	Q9AKH7	Q9akh7 rickettsia	189	40.5	9.6	86	15	Q38258	Q38258 human immun
117	42	9.9	89	10	Q8RX10	Q8rx10 arabidopsis	190	40.5	9.6	86	15	Q38259	Q38259 human immun
118	42	9.9	89	16	Q9K781	Q9k781 bacillus ha	191	40.5	9.6	86	15	Q38260	Q38260 human immun
119	42	9.9	90	2	P72094	P72094 neisseria m	192	40.5	9.6	86	15	Q38261	Q38261 human immun
120	42	9.9	90	15	Q41594	Q41594 human immun	193	40.5	9.6	86	15	Q38262	Q38262 human immun
121	41.5	9.8	63	11	Q9CT15	Q9ct15 mus musculus	194	40.5	9.6	86	15	Q38263	Q38263 human immun
122	41.5	9.8	66	5	Q9VE14	Q9ve14 drosophila	195	40.5	9.6	86	15	Q38264	Q38264 human immun
123	41.5	9.8	66	15	Q9J127	Q9j127 human immun	196	40.5	9.6	86	15	Q38265	Q38265 human immun
124	41.5	9.8	70	14	Q9J178	Q9j178 uncultured	197	40.5	9.6	86	15	Q38267	Q38267 human immun
125	41.5	9.8	70	15	Q97602	Q97602 human immun	198	40.5	9.6	86	15	Q38268	Q38268 human immun
126	41.5	9.8	72	15	Q97H04	Q97hu4 clostridium	199	40.5	9.6	86	15	Q38271	Q38271 human immun
127	41.5	9.8	66	15	Q98273	Q98273 human immun	200	40.5	9.6	86	15	Q38274	Q38274 human immun
128	41.5	9.8	66	15	Q982B4	Q982b4 rhizobium l	201	40.5	9.6	86	15	Q38275	Q38275 human immun
129	41.5	9.8	88	5	Q9NG29	Q9ng29 leishmania	202	40.5	9.6	86	15	Q38276	Q38276 human immun
130	41.5	9.8	89	10	Q9LGM8	Q9lgm8 oryza sativ	203	40.5	9.6	86	15	Q38277	Q38277 human immun
131	41.5	9.8	89	16	Q9J3J7	Q9j3j7 streptomyce	204	40.5	9.6	86	15	Q38278	Q38278 human immun
132	41.5	9.8	63	9	Q9S060	Q9s060 mycobacteri	205	40.5	9.6	86	15	Q38279	Q38279 human immun
133	41.5	9.7	63	10	Q9M590	Q9m590 nicotiana t	206	40.5	9.6	86	15	Q38280	Q38280 human immun
134	41.5	9.7	65	10	Q80566	Q80566 zea mays (m	207	40.5	9.6	86	15	Q38281	Q38281 human immun
135	41.5	9.7	65	15	Q9J115	Q9j115 human immun	208	40.5	9.6	86	15	Q38283	Q38283 human immun
136	41.5	9.7	66	12	Q9V9T5	Q9v9t5 african swi	209	40.5	9.6	86	15	Q38284	Q38284 human immun
137	41.5	9.7	66	15	Q97598	Q97598 human immun	210	40.5	9.6	86	15	Q38287	Q38287 human immun
138	41.5	9.7	67	15	Q9J141	Q9j141 human immun	211	40.5	9.6	86	15	Q38288	Q38288 human immun
139	41.5	9.7	67	15	Q97590	Q97590 human immun	212	40.5	9.6	86	15	Q38289	Q38289 human immun
140	41.5	9.7	68	2	Q46605	Q46605 desulfovibr	213	40.5	9.6	86	15	Q38290	Q38290 human immun
141	41.5	9.7	69	10	Q41693	Q41693 vigna radia	214	40.5	9.6	86	15	Q38293	Q38293 human immun
142	41.5	9.7	70	5	Q9VV37	Q9vv37 drosophila	215	40.5	9.6	86	15	Q38296	Q38296 human immun
143	41.5	9.7	71	12	Q91ET4	Q91et4 cydia pomon	216	40.5	9.6	86	15	Q38297	Q38297 human immun
144	41.5	9.7	71	15	Q97585	Q97585 human immun	217	40.5	9.6	86	15	Q38298	Q38298 human immun
145	41.5	9.7	72	15	Q97591	Q97591 human immun	218	40.5	9.6	86	15	Q38301	Q38301 human immun
146	41.5	9.7	75	15	Q76056	Q76056 human immun	219	40.5	9.6	86	15	Q38302	Q38302 human immun
147	41.5	9.7	76	6	Q8WMJ9	Q8wmj9 macaca mula	220	40.5	9.6	86	15	Q38303	Q38303 human immun
148	41.5	9.7	78	12	Q9PX51	Q9px51 shopee fibro	221	40.5	9.6	86	15	Q38304	Q38304 human immun
149	41.5	9.7	79	4	Q9XRP2	Q9xrp2 homo sapien	222	40.5	9.6	86	15	Q38305	Q38305 human immun
150	41.5	9.7	85	2	Q9WKG1	Q9wkg1 xanthomonas	223	40.5	9.6	86	15	Q38306	Q38306 human immun
151	41.5	9.7	86	4	Q9H323	Q9h323 homo sapien	224	40.5	9.6	86	15	Q38307	Q38307 human immun
152	41.5	9.7	86	16	Q9JZ77	Q9jzt7 neisseria m	225	40.5	9.6	86	15	Q38308	Q38308 human immun
153	41.5	9.7	87	17	Q9TRK4	Q8tkz4 methanosarc	226	40.5	9.6	86	15	Q38309	Q38309 human immun
154	41.5	9.7	88	10	Q42092	Q42092 arabidopsis	227	40.5	9.6	86	15	Q38310	Q38310 human immun
155	41.5	9.7	90	2	Q9AM06	Q9am06 plectonema	228	40.5	9.6	86	15	Q38311	Q38311 human immun
156	40.5	9.6	52	4	Q9GGJ3	Q9ggj3 homo sapien	229	40.5	9.6	86	15	Q38312	Q38312 human immun
157	40.5	9.6	66	15	Q9J186	Q9j186 human immun	230	40.5	9.6	86	15	Q38313	Q38313 human immun
158	40.5	9.6	66	15	Q38432	Q38432 human immun	231	40.5	9.6	86	15	Q38314	Q38314 human immun
159	40.5	9.6	67	15	Q9J176	Q9j176 human immun	232	40.5	9.6	86	15	Q38315	Q38315 human immun
160	40.5	9.6	68	2	Q50886	Q50886 myxococcus	233	40.5	9.6	86	15	Q38316	Q38316 human immun
161	40.5	9.6	68	15	Q74620	Q74620 human immun	234	40.5	9.6	86	15	Q38317	Q38317 human immun
162	40.5	9.6	68	15	Q74630	Q74630 human immun	235	40.5	9.6	86	15	Q38319	Q38319 human immun

236	40.5	9.6	86	15	038320	human	immun	038320	309	40	9.5	67	15	097599	human	immun
237	40.5	9.6	86	15	038321	human	immun	038321	310	40	9.5	67	15	097607	human	immun
238	40.5	9.6	86	15	038322	human	immun	038322	311	40	9.5	67	16	097157	clostridium	
239	40.5	9.6	86	15	038323	human	immun	038323	312	40	9.5	68	15	097196	human	immun
240	40.5	9.6	86	15	038324	human	immun	038324	313	40	9.5	68	15	097191	human	immun
241	40.5	9.6	86	15	038325	human	immun	038325	314	40	9.5	68	15	097180	human	immun
242	40.5	9.6	86	15	038326	human	immun	038326	315	40	9.5	68	15	097179	human	immun
243	40.5	9.6	86	15	038327	human	immun	038327	316	40	9.5	68	15	097165	human	immun
244	40.5	9.6	86	15	038328	human	immun	038328	317	40	9.5	68	15	097164	human	immun
245	40.5	9.6	86	15	038330	human	immun	038330	318	40	9.5	68	15	097164	human	immun
246	40.5	9.6	86	15	038331	human	immun	038331	319	40	9.5	68	15	097146	human	immun
247	40.5	9.6	86	15	038333	human	immun	038333	320	40	9.5	68	15	097057	human	immun
248	40.5	9.6	86	15	038334	human	immun	038334	321	40	9.5	69	15	097134	human	immun
249	40.5	9.6	86	15	038451	human	immun	038451	322	40	9.5	69	15	097119	human	immun
250	40.5	9.6	86	15	038452	human	immun	038452	323	40	9.5	70	5	094857	drosohilla	
251	40.5	9.6	86	15	038453	human	immun	038453	324	40	9.5	70	5	094871	drosohilla	
252	40.5	9.6	86	15	038455	human	immun	038455	325	40	9.5	70	5	094937	drosohilla	
253	40.5	9.6	86	15	038458	human	immun	038458	326	40	9.5	70	5	094951	drosohilla	
254	40.5	9.6	86	15	038536	human	immun	038536	327	40	9.5	70	5	094951	drosohilla	
255	40.5	9.6	88	15	038536	human	immun	038536	328	40	9.5	70	15	097623	human	immun
256	40.5	9.6	88	15	038539	human	immun	038539	329	40	9.5	71	15	097611	human	immun
257	40.5	9.6	89	15	097228	human	immun	097228	330	40	9.5	71	15	097157	human	immun
258	40.5	9.6	89	15	079821	human	immun	079821	331	40	9.5	72	15	0976045	human	immun
259	40.5	9.6	89	15	038050	human	immun	038050	332	40	9.5	72	15	0976045	human	immun
260	40.5	9.6	89	15	038063	human	immun	038063	333	40	9.5	73	3	092288	coctidolo	
261	40.5	9.6	89	15	038064	human	immun	038064	334	40	9.5	73	3	092288	coctidolo	
262	40.5	9.6	89	15	038346	human	immun	038346	335	40	9.5	73	15	097606	human	immun
263	40.5	9.6	89	15	038347	human	immun	038347	336	40	9.5	74	15	097606	human	immun
264	40.5	9.6	89	15	038348	human	immun	038348	337	40	9.5	74	15	097606	human	immun
265	40.5	9.6	89	15	038350	human	immun	038350	338	40	9.5	74	15	097606	human	immun
266	40.5	9.6	90	4	09B0L1	homo sapien		09B0L1	339	40	9.5	74	15	097606	human	immun
267	40.5	9.6	90	10	09XHD5	ipomoea bat		09XHD5	340	40	9.5	74	15	097606	human	immun
268	40	9.5	42	15	097043	mus musculus		097043	341	40	9.5	75	15	097606	human	immun
269	40	9.5	44	15	09J193	human	immun	09J193	342	40	9.5	75	15	097606	human	immun
270	40	9.5	47	15	09J188	human	immun	09J188	343	40	9.5	75	15	097606	human	immun
271	40	9.5	48	15	09J161	human	immun	09J161	344	40	9.5	75	15	097606	human	immun
272	40	9.5	47	15	097604	human	immun	097604	345	40	9.5	75	15	097606	human	immun
273	40	9.5	51	16	09JXZ6	neisseria m		09JXZ6	346	40	9.5	75	15	097606	human	immun
274	40	9.5	53	15	087381	chimpanzee		087381	347	40	9.5	76	12	09JYR8	human	immun
275	40	9.5	54	15	087411	chimpanzee		087411	348	40	9.5	76	12	09JYR8	human	immun
276	40	9.5	56	15	087417	chimpanzee		087417	349	40	9.5	76	15	097606	human	immun
277	40	9.5	58	15	09J133	human	immun	09J133	350	40	9.5	76	15	097606	human	immun
278	40	9.5	59	15	097600	human	immun	097600	351	40	9.5	76	15	097606	human	immun
279	40	9.5	59	15	097606	human	immun	097606	352	40	9.5	77	2	09X5N9	human	immun
280	40	9.5	60	15	094844	drosohilla		094844	353	40	9.5	77	15	097594	human	immun
281	40	9.5	60	11	09J182	mus musculus		09J182	354	40	9.5	78	10	09AY11	guillardia	
282	40	9.5	60	11	08E556	mus musculus		08E556	355	40	9.5	78	15	097605	human	immun
283	40	9.5	60	15	094326	human	immun	040336	356	40	9.5	81	16	08S2K8	oryza sativ	
284	40	9.5	60	15	097586	human	immun	097586	357	40	9.5	81	16	08S2K8	rhizobium m	
285	40	9.5	61	2	09BVR8	streptococc		09BVR8	358	40	9.5	82	10	080943	arabidopsis	
286	40	9.5	61	15	09J198	human	immun	09J198	359	40	9.5	83	10	094JCA	oryza sativ	
287	40	9.5	61	15	096047	human	immun	096047	360	40	9.5	83	10	09KKS0	vibrio chol	
288	40	9.5	62	15	09J132	human	immun	09J132	361	40	9.5	84	12	09WKH1	encephalomy	
289	40	9.5	62	15	097612	human	immun	097612	362	40	9.5	85	4	09BVM6	homo sapien	
290	40	9.5	63	15	09J177	human	immun	09J177	363	40	9.5	85	11	09ER98	mus musculus	
291	40	9.5	63	16	08YK99	human	immun	08YK99	364	40	9.5	86	15	038251	human	immun
292	40	9.5	64	15	09J192	anabaena sp		09J192	365	40	9.5	86	15	038272	human	immun
293	40	9.5	64	15	09J189	human	immun	09J189	366	40	9.5	86	15	038282	human	immun
294	40	9.5	64	15	097601	human	immun	097601	367	40	9.5	86	15	038332	human	immun
295	40	9.5	64	15	075980	human	immun	075980	368	40	9.5	86	15	038460	human	immun
296	40	9.5	65	15	09J195	human	immun	09J195	369	40	9.5	87	12	09DL09	human	immun
297	40	9.5	65	15	09J137	human	immun	09J137	370	40	9.5	87	15	038510	zucchini ye	
298	40	9.5	66	15	09J173	human	immun	09J173	371	40	9.5	87	15	038514	human	immun
299	40	9.5	66	15	09J154	human	immun	09J154	372	40	9.5	88	16	073494	synechocyst	
300	40	9.5	66	15	097597	human	immun	097597	373	40	9.5	88	16	099204	streptococc	
301	40	9.5	66	15	076266	human	immun	076266	374	40	9.5	89	15	038344	human	immun
302	40	9.5	67	15	09J187	human	immun	09J187	375	40	9.5	89	15	038345	human	immun
303	40	9.5	67	15	09J180	human	immun	09J180	376	40	9.5	90	15	079820	human	immun
304	40	9.5	67	15	09J126	human	immun	09J126	377	40	9.5	90	16	096700	human	immun
305	40	9.5	67	15	09J118	human	immun	09J118	378	39.5	9.3	56	16	09TMS1	bacillus su	
306	40	9.5	67	15	09J116	human	immun	09J116	379	39.5	9.3	60	9	027193	Qymal1 anabaena sp	
307	40	9.5	67	15	09J116	human	immun	09J116	380	39.5	9.3	60	9	09MC67	tetrahymena	
308	40	9.5	67	15	097593	human	immun	097593	381	39.5	9.3	61	16	09CHZ4	bacterioph	
												65	15	09J120	lactococcus	
															human	immun



382	39.5	9.3	66	12	Q91JB7	Q91JB7 hepatitis c	455	39	9.2	90	16	Q92854	Q92854 chlamydia p
383	39.5	9.3	67	16	Q92EM1	Q92EM1 listeria in	456	38.5	9.1	51	18	Q9X7N4	Q9X7N4 streptomyce
384	39.5	9.3	67	16	Q94M8	Q94M8 listeria mo	457	38.5	9.1	48	2	Q93MX5	Q93MX5 uncultured
385	39.5	9.3	67	16	Q94M8	Q94M8 streptomyce	458	38.5	9.1	53	10	Q9XG2	Q9XG2 arabidopsis
386	39.5	9.3	68	15	Q96G3	Q96G3 human immun	459	38.5	9.1	60	4	Q94F1	Q94F1 homo sapien
387	39.5	9.3	68	15	Q9465	Q9465 human immun	460	38.5	9.1	60	16	Q9UAC2	Q9UAC2 agrobacteri
388	39.5	9.3	69	15	Q9WKR3	Q9WKR3 human immun	461	38.5	9.1	63	2	Q9EHM8	Q9EHM8 comamonas t
389	39.5	9.3	72	16	Q92IK5	Q92IK5 rickettsia	462	38.5	9.1	63	2	Q9S852	Q9S852 sphingomona
390	39.5	9.3	73	4	Q93067	Q93067 homo sapien	463	38.5	9.1	64	16	Q9K0E5	Q9K0E5 neisseria m
391	39.5	9.3	73	4	Q98ZL1	Q98ZL1 homo sapien	464	38.5	9.1	66	2	Q9ZEH6	Q9ZEH6 enterococcu
392	39.5	9.3	73	11	Q9EPV8	Q9EPV8 mus musculus	465	38.5	9.1	66	6	Q9WKK0	Q9WKK0 macaca mula
393	39.5	9.3	75	16	Q9JV21	Q9JV21 neisseria m	466	38.5	9.1	67	16	Q9ZK90	Q9ZK90 listeria mo
394	39.5	9.3	76	2	Q9AM01	Q9AM01 rhodoferrax	467	38.5	9.1	68	15	Q9Y9V4	Q9Y9V4 human immun
395	39.5	9.3	76	2	Q9AM01	Q9AM01 helicobacte	468	38.5	9.1	69	2	Q91902	Q91902 proteus mir
396	39.5	9.3	77	16	Q97TB3	Q97TB3 streptococ	469	38.5	9.1	69	11	Q9CWL5	Q9CWL5 mus musculu
397	39.5	9.3	77	16	Q9YXU8	Q9YXU8 anabaena sp	470	38.5	9.1	69	11	Q9WMT2	Q9WMT2 human immun
398	39.5	9.3	80	15	Q9GST4	Q9GST4 human immun	471	38.5	9.1	69	15	Q9AFY0	Q9AFY0 shigella fl
399	39.5	9.3	81	17	Q97V52	Q97V52 sulfolobus	472	38.5	9.1	74	2	Q9VN30	Q9VN30 helicobacte
400	39.5	9.3	82	10	Q9LNN9	Q9LNN9 arabidopsis	473	38.5	9.1	76	16	Q9S747	Q9S747 helicobacte
401	39.5	9.3	86	12	Q9VAA3	Q9VAA3 white spot	474	38.5	9.1	76	16	Q9KQC4	Q9KQC4 vibrio chol
402	39.5	9.3	86	16	Q92K09	Q92K09 rhizobium m	475	38.5	9.1	78	16	Q9XG51	Q9XG51 salmonella
403	39.5	9.3	89	2	Q92972	Q92972 rhizobium m	476	38.5	9.1	80	13	Q90XQ0	Q90XQ0 ambystoma m
404	39.5	9.3	89	10	Q9XES8	Q9XES8 glycine max	477	38.5	9.1	80	14	Q99IV0	Q99IV0 uncultured
405	39.5	9.3	89	10	Q92414	Q92414 oryza sativ	478	38.5	9.1	81	15	Q90Q06	Q90Q06 human immun
406	39.5	9.3	89	15	Q9YV38	Q9YV38 human immun	479	38.5	9.1	82	15	Q9YKZ5	Q9YKZ5 human immun
407	39.5	9.3	89	15	Q9YV35	Q9YV35 human immun	480	38.5	9.1	82	15	Q9YKZ5	Q9YKZ5 human immun
408	39.5	9.3	89	15	Q9YV25	Q9YV25 human immun	481	38.5	9.1	85	2	Q9BHZ6	Q9BHZ6 shevanelia
409	39.5	9.3	89	15	Q9YV24	Q9YV24 human immun	482	38.5	9.1	85	12	Q95693	Q95693 papaya leaf
410	39.5	9.3	89	15	Q94887	Q94887 human immun	483	38.5	9.1	86	10	Q9FU06	Q9FU06 oryza sativ
411	39.5	9.3	89	15	Q9WA05	Q9WA05 human immun	484	38.5	9.1	86	11	Q99LV9	Q99LV9 mus musculu
412	39.5	9.3	89	15	Q9W8F1	Q9W8F1 human immun	485	38.5	9.1	86	11	Q93248	Q93248 human immun
413	39.5	9.3	89	15	Q9YJK2	Q9YJK2 human immun	486	38.5	9.1	86	15	Q98291	Q98291 human immun
414	39.5	9.3	89	15	Q9YJ41	Q9YJ41 human immun	487	38.5	9.1	86	15	Q98291	Q98291 meleagris g
415	39.5	9.3	89	15	Q9W9S7	Q9W9S7 human immun	488	38.5	9.1	87	13	Q90X98	Q90X98 human immun
416	39.5	9.3	89	15	Q9W8S2	Q9W8S2 human immun	489	38.5	9.1	89	15	Q90533	Q90533 human immun
417	39.5	9.3	90	16	Q95312	Q95312 mycobacteri	490	38.5	9.1	89	15	Q94886	Q94886 human immun
418	39	9.2	41	6	Q918852	Q918852 macaca radi	491	38.5	9.1	89	16	Q9ZCAL	Q9ZCAL listeria in
419	39	9.2	48	11	Q93365	Q93365 rattus norv	492	38.5	9.1	90	16	Q9BZL4	Q9BZL4 salmonella
420	39	9.2	51	4	Q9H3T1	Q9H3T1 homo sapien	493	38.5	9.1	90	16	Q9BZL4	Q9BZL4 comastoma c
421	39	9.2	53	5	Q927909	Q927909 chimpanzee	494	38.5	9.1	90	16	Q91557	Q91557 human immun
422	39	9.2	53	5	Q927909	Q927909 chimpanzee	495	38.5	9.1	90	16	Q91557	Q91557 thermoanaer
423	39	9.2	53	5	Q927909	Q927909 chimpanzee	496	38	9.0	44	6	Q9TTF8	Q9TTF8 ateles belz
424	39	9.2	55	12	Q96728	Q96728 equid herpe	497	38	9.0	53	15	Q97357	Q97357 chimpanzee
425	39	9.2	55	12	Q96728	Q96728 equid herpe	498	38	9.0	55	2	Q45363	Q45363 bordetella
426	39	9.2	58	15	Q97330	Q97330 chimpanzee	499	38	9.0	55	12	Q9WPG9	Q9WPG9 zebra herpe
427	39	9.2	61	17	Q972E5	Q972E5 sulfolobus	500	38	9.0	56	16	Q92KY3	Q92KY3 rhizobium m
428	39	9.2	62	10	Q970H2	Q970H2 arabidopsis	501	38	9.0	60	10	Q9W2K3	Q9W2K3 arabidopsis
429	39	9.2	64	11	Q92675	Q92675 rattus norv	502	38	9.0	61	2	Q9AML5	Q9AML5 edwardsiell
430	39	9.2	67	4	Q9H13	Q9H13 gallus gall	503	38	9.0	61	5	Q92565	Q92565 procambatus
431	39	9.2	68	4	Q9N143	Q9N143 homo sapien	504	38	9.0	62	4	Q9WVQ4	Q9WVQ4 homo sapien
432	39	9.2	68	6	Q91320	Q91320 bos taurus	505	38	9.0	62	6	Q9TSE1	Q9TSE1 sus scrofa
433	39	9.2	69	5	Q9T2B4	Q9T2B4 dictyostell	506	38	9.0	62	16	Q9YTW2	Q9YTW2 anabaena sp
434	39	9.2	70	11	Q90X17	Q90X17 rattus norv	507	38	9.0	63	2	Q9AME4	Q9AME4 desulfohalob
435	39	9.2	71	15	Q90586	Q90586 human immun	508	38	9.0	63	5	Q91635	Q91635 drosophila
436	39	9.2	72	15	Q9Y135	Q9Y135 human immun	509	38	9.0	64	16	Q91982	Q91982 sus scrofa
437	39	9.2	72	15	Q91F20	Q91F20 human immun	510	38	9.0	64	16	Q91982	Q91982 sus scrofa
438	39	9.2	73	16	Q98YD3	Q98YD3 brucella me	511	38	9.0	64	16	Q92X59	Q92X59 xylella fas
439	39	9.2	73	12	Q9YPL4	Q9YPL4 encephalomy	512	38	9.0	64	16	Q92X59	Q92X59 rhizobium m
440	39	9.2	76	12	Q94948	Q94948 avian infec	513	38	9.0	64	16	Q92X59	Q92X59 rickettsia
441	39	9.2	77	6	Q97679	Q97679 trichosurus	514	38	9.0	67	2	Q9J160	Q9J160 human immun
442	39	9.2	77	16	Q9HZJ7	Q9HZJ7 pseudomonas	515	38	9.0	67	12	Q9F681	Q9F681 streptococc
443	39	9.2	77	16	Q9K4K2	Q9K4K2 streptomyce	516	38	9.0	69	10	Q920745	Q920745 encephalomy
444	39	9.2	80	3	Q9AFQ4	Q9AFQ4 shigella fl	517	38	9.0	69	12	Q92562	Q92562 elaeagnus u
445	39	9.2	80	3	Q9HGR5	Q9HGR5 mucor pirif	518	38	9.0	69	12	Q9QF53	Q9QF53 hantavirus
446	39	9.2	80	12	Q91726	Q91726 grapevine r	519	38	9.0	70	2	Q9AFK7	Q9AFK7 shigella fl
447	39	9.2	80	12	Q91903	Q91903 rupestris s	520	38	9.0	71	15	Q97589	Q97589 human immun
448	39	9.2	81	16	Q92HE6	Q92HE6 rickettsia	521	38	9.0	72	4	Q14813	Q14813 homo sapien
449	39	9.2	81	16	Q97770	Q97770 mycobacteri	522	38	9.0	74	2	Q9X7D4	Q9X7D4 mycobacteri
450	39	9.2	82	5	Q9U9M2	Q9U9M2 hirtodrosop	523	38	9.0	74	4	Q9BTF1	Q9BTF1 homo sapien
451	39	9.2	82	5	Q9U9M2	Q9U9M2 hirtodrosop	524	38	9.0	74	5	Q9XXH1	Q9XXH1 caenorhabdi
452	39	9.2	87	10	Q96743	Q96743 arabidopsis	525	38	9.0	75	6	Q9MZA8	Q9MZA8 sus scrofa
453	39	9.2	88	10	Q98537	Q98537 human immun	526	38	9.0	75	6	Q92U21	Q92U21 rhizobium m
454	39	9.2	88	10	Q9850W	Q9850W oryza sativ	527	38	9.0	76	12	Q64946	Q64946 avian infec

528	9.0	76	12	Q64950	Q64950 avian infec	601	37.5	8.9	68	15	Q74261	Q74261 human immun	
529	38	9.0	76	12	Q09529	Q09529 human immun	602	37.5	8.9	68	15	Q74263	Q74263 human immun
530	38	9.0	78	12	Q40608	Q40608 hepatitis c	603	37.5	8.9	68	15	Q74265	Q74265 human immun
531	38	9.0	78	12	Q40609	Q40609 hepatitis c	604	37.5	8.9	68	15	Q74267	Q74267 human immun
532	38	9.0	79	2	Q924G7	Q924G7 escherichia	605	37.5	8.9	68	15	Q74269	Q74269 human immun
533	38	9.0	80	6	Q9MSR2	Q9MSR2 sus scrofa	606	37.5	8.9	68	15	Q74271	Q74271 human immun
534	38	9.0	80	10	Q9RTS5	Q9RTS5 zea mays (m	607	37.5	8.9	69	4	Q9BU94	Q9BU94 homo sapien
535	38	9.0	81	16	Q8S0N7	Q8S0N7 oryza sativ	608	37.5	8.9	69	12	Q8QMG3	Q8QMG3 human coxxa
536	38	9.0	81	16	Q8S0N7	Q8S0N7 oryza sativ	609	37.5	8.9	69	12	Q8QMG3	Q8QMG3 human coxxa
537	38	9.0	82	5	Q92Z24	Q92Z24 streptomyce	610	37.5	8.9	69	15	Q9WMU2	Q9WMU2 human immun
538	38	9.0	82	5	Q92Z24	Q92Z24 rhizobium m	611	37.5	8.9	69	15	Q9WMT1	Q9WMT1 human immun
539	38	9.0	85	11	Q18505	Q18505 stronglyloid	612	37.5	8.9	69	15	Q9WMT1	Q9WMT1 human immun
540	38	9.0	85	11	Q8VDG4	Q8VDG4 mus musculus	613	37.5	8.9	71	1	Q977L2	Q977L2 uncultured
541	38	9.0	85	16	Q98JW6	Q98JW6 rhizobium l	614	37.5	8.9	72	15	Q9WFE1	Q9WFE1 human immun
542	38	9.0	86	9	Q9L5F4	Q9L5F4 salmonella	615	37.5	8.9	74	5	Q9NLX4	Q9NLX4 leishmania
543	38	9.0	88	6	Q62851	Q62851 mycobacteri	616	37.5	8.9	74	5	Q9NLX4	Q9NLX4 leishmania
544	38	9.0	90	10	Q94HS2	Q94HS2 ovis aries	617	37.5	8.9	74	10	Q9LX2P5	Q9LX2P5 arabidopsis
545	38	9.0	90	10	Q94HS2	Q94HS2 ovis aries	618	37.5	8.9	74	10	Q9LX2P5	Q9LX2P5 arabidopsis
546	37.5	8.9	47	15	Q73724	Q73724 human immun	619	37.5	8.9	74	15	Q9IF12	Q9IF12 human immun
547	37.5	8.9	47	15	Q73726	Q73726 human immun	620	37.5	8.9	74	16	Q53491	Q53491 mycobacteri
548	37.5	8.9	47	15	Q73728	Q73728 human immun	621	37.5	8.9	76	2	Q8VN32	Q8VN32 helicobacte
549	37.5	8.9	47	15	Q73729	Q73729 human immun	622	37.5	8.9	77	6	Q9GMR8	Q9GMR8 macaca fasc
550	37.5	8.9	47	15	Q73731	Q73731 human immun	623	37.5	8.9	77	10	Q9KIX4	Q9KIX4 oryza sativ
551	37.5	8.9	47	15	Q73733	Q73733 human immun	624	37.5	8.9	78	15	Q8RTG1	Q8RTG1 human immun
552	37.5	8.9	47	15	Q73733	Q73733 human immun	625	37.5	8.9	78	15	Q8RTG1	Q8RTG1 human immun
553	37.5	8.9	47	15	Q73763	Q73763 human immun	626	37.5	8.9	78	15	Q8RTG1	Q8RTG1 human immun
554	37.5	8.9	47	15	Q73767	Q73767 human immun	627	37.5	8.9	78	15	Q8RTG1	Q8RTG1 human immun
555	37.5	8.9	47	15	Q73769	Q73769 human immun	628	37.5	8.9	78	15	Q8RTG1	Q8RTG1 human immun
556	37.5	8.9	47	15	Q73771	Q73771 human immun	629	37.5	8.9	78	15	Q8RTG1	Q8RTG1 human immun
557	37.5	8.9	47	15	Q73773	Q73773 human immun	630	37.5	8.9	78	15	Q8RTG1	Q8RTG1 human immun
558	37.5	8.9	49	15	Q73811	Q73811 human immun	631	37.5	8.9	78	15	Q8RTG1	Q8RTG1 human immun
559	37.5	8.9	49	15	Q73813	Q73813 human immun	632	37.5					

674	37	8.7	42	2	P82555 streptococc	747	37	8.7	79	15	038467	038467 human immun
675	37	8.7	44	15	O89363 human immun	748	37	8.7	79	15	038471	038471 human immun
676	37	8.7	45	2	Q47202 escherichia	749	37	8.7	79	16	Q9PHI2	Q9PHI2 xylella fas
677	37	8.7	45	12	Q9DS22 salmireline	750	37	8.7	80	2	Q51000	Q51000 neisseria g
678	37	8.7	50	2	Q9RAD1 rhodofexer	751	37	8.7	80	5	Q8STA4	Q8STA4 penaeus jap
679	37	8.7	50	16	Q98D00 rhizobium l	752	37	8.7	80	9	O64261	O64261 mycobacteri
680	37	8.7	52	16	Q92NN7 rhizobium m	753	37	8.7	80	15	038439	038439 human immun
681	37	8.7	53	2	Q9RL85 staphylococ	754	37	8.7	81	10	O04814	O04814 sporobolus
682	37	8.7	53	5	Q9U7L0 plasmodium	755	37	8.7	81	12	Q9YPK9	Q9YPK9 encephalomy
683	37	8.7	53	15	Q87399 chimpanzee	756	37	8.7	81	12	O8VAP4	O8VAP4 white spot
684	37	8.7	53	15	O87405 chimpanzee	757	37	8.7	81	12	O9W8J1	O9W8J1 encephalomy
685	37	8.7	53	15	O87414 chimpanzee	758	37	8.7	82	10	O9FP81	O9FP81 oryza sativ
686	37	8.7	53	16	O8U422 agrobacteri	759	37	8.7	83	5	O8T960	O8T960 drosophila
687	37	8.7	53	17	O96XV5 sulfolobus	760	37	8.7	83	12	Q9YPK3	Q9YPK3 porcine enc
688	37	8.7	57	5	O8T7Y9 caenorhabdi	761	37	8.7	83	12	O9VPL3	O9VPL3 encephalomy
689	37	8.7	58	15	O87420 chimpanzee	762	37	8.7	83	12	Q9WK36	Q9WK36 encephalomy
690	37	8.7	59	12	Q39817 encephalomy	763	37	8.7	84	10	O9FPA8	O9FPA8 oryza sativ
691	37	8.7	62	7	O98149 papio hamad	764	37	8.7	84	16	P74471	P74471 synechocyst
692	37	8.7	62	7	Q9XRX3 papio cynoc	765	37	8.7	84	16	O55568	O55568 synechocyst
693	37	8.7	62	7	Q30881 papio hamad	766	37	8.7	84	16	O9KAV4	O9KAV4 bacillus ha
694	37	8.7	62	7	Q30591 macaca mula	767	37	8.7	84	16	O8VI90	O8VI90 streptomyce
695	37	8.7	62	15	O87395 chimpanzee	768	37	8.7	84	16	O9XAG7	O9XAG7 streptomyce
696	37	8.7	62	16	O9PHB7 xylella fas	769	37	8.7	85	5	O9VDM5	O9VDM5 drosophila
697	37	8.7	63	2	Q92431 pseudomonas	770	37	8.7	85	6	O13114	O13114 isodon mac
698	37	8.7	63	2	Q9R3T3 pseudomonas	771	37	8.7	85	6	O13122	O13122 tarsipes ro
699	37	8.7	63	11	Q99MY4 mus musculu	772	37	8.7	85	6	O02795	O02795 ornithorhyn
700	37	8.7	64	15	Q9J167 human immun	773	37	8.7	85	6	O02798	O02798 petaurus br
701	37	8.7	64	15	Q9J155 human immun	774	37	8.7	85	6	O13104	O13104 cercarctus
702	37	8.7	64	16	Q8UCK5 agrobacteri	775	37	8.7	85	6	O02803	O02803 trichosurus
703	37	8.7	65	2	O07265 mycobacteri	776	37	8.7	85	6	O02792	O02792 notoryctes
704	37	8.7	65	15	Q9J147 human immun	777	37	8.7	85	6	O02790	O02790 macropus ru
705	37	8.7	65	15	Q9J139 human immun	778	37	8.7	85	6	O13105	O13105 dasyuroides
706	37	8.7	65	15	Q97584 human immun	779	37	8.7	85	6	O02801	O02801 tachyglossu
707	37	8.7	66	15	Q9J185 human immun	780	37	8.7	86	11	Q9JIK6	Q9JIK6 mus musculu
708	37	8.7	66	16	Q8ZCY3 yersinia pe	781	37	8.7	86	15	O38459	O38459 human immun
709	37	8.7	66	17	Q9VEZ2 aeropyrum p	782	37	8.7	86	15	O38425	O38425 human immun
710	37	8.7	67	15	Q97609 human immun	783	37	8.7	87	15	O38426	O38426 human immun
711	37	8.7	67	15	O38429 human immun	784	37	8.7	87	15	O38427	O38427 human immun
712	37	8.7	67	15	O38431 human immun	785	37	8.7	87	15	O38444	O38444 human immun
713	37	8.7	67	15	Q9J166 human immun	786	37	8.7	87	15	O38445	O38445 human immun
714	37	8.7	68	15	Q97603 human immun	787	37	8.7	87	15	O38447	O38447 human immun
715	37	8.7	68	15	Q38430 human immun	788	37	8.7	87	15	O38465	O38465 human immun
716	37	8.7	69	12	O8QMD8 human echov	789	37	8.7	87	15	O38470	O38470 human immun
717	37	8.7	69	15	Q90540 human immun	790	37	8.7	87	15	O38473	O38473 human immun
718	37	8.7	69	15	Q97613 human immun	791	37	8.7	87	15	O38474	O38474 human immun
719	37	8.7	71	10	Q9XGY3 malus domes	792	37	8.7	87	15	O38475	O38475 human immun
720	37	8.7	71	10	Q90584 human immun	793	37	8.7	87	15	O38478	O38478 human immun
721	37	8.7	72	11	Q9LWE4 mus musculu	794	37	8.7	87	15	O38479	O38479 human immun
722	37	8.7	72	17	Q8Z557 Proboscium	795	37	8.7	87	15	O38480	O38480 human immun
723	37	8.7	72	17	Q9L196 arbidopsia	796	37	8.7	87	15	O38481	O38481 human immun
724	37	8.7	73	11	Q8R4G4 rattus norv	797	37	8.7	87	15	O38484	O38484 human immun
725	37	8.7	73	11	O8R1H0 mus musculu	798	37	8.7	87	15	O38485	O38485 human immun
726	37	8.7	73	15	Q9J170 human immun	799	37	8.7	87	15	O38486	O38486 human immun
727	37	8.7	73	15	O38428 human immun	800	37	8.7	87	15	O38488	O38488 human immun
728	37	8.7	73	15	O38428 human immun	801	37	8.7	87	15	O38494	O38494 human immun
729	37	8.7	73	16	Q8ZEP2 yersinia pe	802	37	8.7	87	15	O38496	O38496 human immun
730	37	8.7	74	8	Q35301 oryza sativ	803	37	8.7	87	15	O38500	O38500 human immun
731	37	8.7	74	12	Q9W8B2 encephalomy	804	37	8.7	87	15	O38502	O38502 human immun
732	37	8.7	75	6	Q95LM4 macaca fasc	805	37	8.7	87	15	O38503	O38503 human immun
733	37	8.7	75	11	Q8R043 mus musculu	806	37	8.7	87	15	O38504	O38504 human immun
734	37	8.7	75	12	Q9WK35 encephalomy	807	37	8.7	87	15	O38507	O38507 human immun
735	37	8.7	75	15	O38438 human immun	808	37	8.7	87	15	O38512	O38512 human immun
736	37	8.7	75	15	O38491 human immun	809	37	8.7	87	15	O38513	O38513 human immun
737	37	8.7	76	12	Q64943 avian infec	810	37	8.7	87	15	O38515	O38515 human immun
738	37	8.7	76	12	Q68828 human cytom	811	37	8.7	87	15	O38516	O38516 human immun
739	37	8.7	78	2	Q9AF88 anaplasma m	812	37	8.7	87	15	O38518	O38518 human immun
740	37	8.7	78	15	O38433 human immun	813	37	8.7	87	15	O38528	O38528 human immun
741	37	8.7	78	15	O38434 human immun	814	37	8.7	87	15	O38538	O38538 human immun
742	37	8.7	78	15	O38437 human immun	815	37	8.7	87	16	O92YI6	O92YI6 rhizobium m
743	37	8.7	78	15	O38449 human immun	816	37	8.7	87	16	O92YI6	O92YI6 rhizobium m
744	37	8.7	78	15	O38468 human immun	817	37	8.7	88	5	Q95R57	Q95R57 drosophila
745	37	8.7	78	15	O38469 human immun	818	37	8.7	88	5	O9W109	O9W109 drosophila
746	37	8.7	78	16	Q9PBM4 xylella fas	819	37	8.7	88	15	O38476	O38476 human immun

820	37	8.7	88	15	Q38487	O38487 human immun	893	36.5	8.6	85	9	Q9A204	O9A204 bacterioph
821	37	8.7	88	15	Q38495	O38495 human immun	894	36.5	8.6	86	15	Q90V10	Q90V10 human immun
822	37	8.7	88	15	Q38498	O38498 human immun	895	36.5	8.6	86	15	Q38533	O38533 human immun
823	37	8.7	88	15	Q38506	O38506 human immun	896	36.5	8.6	86	16	Q9CHT9	Q9CHT9 lactococcus
824	37	8.7	88	16	Q9P17	Q9P17 xylella fas	897	36.5	8.6	87	2	Q9F125	Q9F125 bacillus ci
825	37	8.7	88	16	Q9A679	Q9A679 thermococ	898	36.5	8.6	87	2	Q9F125	Q9F125 pseudomonas
826	37	8.7	88	17	Q800X6	Q800X6 pyrococcus	899	36.5	8.6	87	2	Q9B311	Q9B311 homo sapien
827	37	8.7	89	2	Q9A016	Q9A016 pseudomonas	900	36.5	8.6	87	4	Q9A080	Q9A080 caulobacter
828	37	8.7	89	16	Q98H18	Q98H18 rhizobium l	901	36.5	8.6	88	2	Q52509	Q52509 pseudomonas
829	37	8.7	90	16	Q922N2	Q922N2 rhizobium m	902	36.5	8.6	88	4	Q9UDK0	Q9UDK0 homo sapien
830	37	8.7	90	13	Q72819	Q72819 human immun	903	36.5	8.6	88	5	Q9VUN4	Q9VUN4 drosophila
831	37	8.7	90	13	Q35438	Q35438 meriones un	904	36.5	8.6	88	8	Q34242	Q34242 calyptogena
832	36.5	8.6	43	16	Q9CHS4	Q9CHS4 lactococcus	905	36.5	8.6	88	8	Q34244	Q34244 calyptogena
833	36.5	8.6	47	5	Q9VBF3	Q9VBF3 drosophila	906	36.5	8.6	89	15	Q9YV34	Q9YV34 human immun
834	36.5	8.6	50	15	Q89372	Q89372 human immun	907	36.5	8.6	89	15	Q9YV33	Q9YV33 human immun
835	36.5	8.6	53	16	Q54165	Q54165 streptomyce	908	36.5	8.6	89	15	Q9YV32	Q9YV32 human immun
836	36.5	8.6	58	10	Q9X1Y4	Q9X1Y4 oryzae sativ	909	36.5	8.6	89	15	Q9YV31	Q9YV31 human immun
837	36.5	8.6	58	16	Q9KAY9	Q9KAY9 bacillus ha	910	36.5	8.6	89	15	Q9YV29	Q9YV29 human immun
838	36.5	8.6	59	5	Q9NCR6	Q9NCR6 cryptospori	911	36.5	8.6	89	15	Q9YV26	Q9YV26 human immun
839	36.5	8.6	61	16	Q8VKNO	Q8VKNO mycobacteri	912	36.5	8.6	89	15	Q38040	Q38040 human immun
840	36.5	8.6	62	4	Q96DC3	Q96DC3 homo sapien	913	36.5	8.6	89	15	Q38041	Q38041 human immun
841	36.5	8.6	62	17	Q73988	Q73988 pyrococcus	914	36.5	8.6	89	15	Q38042	Q38042 human immun
842	36.5	8.6	63	6	Q8S081	Q8S081 pongo pygma	915	36.5	8.6	89	15	Q38043	Q38043 human immun
843	36.5	8.6	63	9	Q9F2V8	Q9F2V8 bacterioph	916	36.5	8.6	89	15	Q38044	Q38044 human immun
844	36.5	8.6	64	13	Q31950	Q31950 gallus gall	917	36.5	8.6	89	15	Q38045	Q38045 human immun
845	36.5	8.6	66	9	Q47322	Q47322 bacterioph	918	36.5	8.6	89	15	Q38046	Q38046 human immun
846	36.5	8.6	66	16	Q8X0U4	Q8X0U4 ralstonia s	919	36.5	8.6	89	15	Q38047	Q38047 human immun
847	36.5	8.6	67	2	Q8VW06	Q8VW06 helicobacte	920	36.5	8.6	89	15	Q38048	Q38048 human immun
848	36.5	8.6	67	2	Q8VW06	Q8VW06 helicobacte	921	36.5	8.6	89	15	Q38049	Q38049 human immun
849	36.5	8.6	68	2	Q8VW06	Q8VW06 helicobacte	922	36.5	8.6	89	15	Q38052	Q38052 human immun
850	36.5	8.6	68	15	Q74306	Q74306 human immun	923	36.5	8.6	89	15	Q38054	Q38054 human immun
851	36.5	8.6	68	16	Q8VJH4	Q8VJH4 mycobacteri	924	36.5	8.6	89	15	Q38055	Q38055 human immun
852	36.5	8.6	69	15	Q9WMU0	Q9WMU0 human immun	925	36.5	8.6	89	15	Q38056	Q38056 human immun
853	36.5	8.6	70	4	Q9H363	Q9H363 homo sapien	926	36.5	8.6	89	15	Q38057	Q38057 human immun
854	36.5	8.6	71	16	Q99SP6	Q99SP6 staphylococ	927	36.5	8.6	89	15	Q38058	Q38058 human immun
855	36.5	8.6	72	3	Q8WZU1	Q8WZU1 neurospora	928	36.5	8.6	89	15	Q38059	Q38059 human immun
856	36.5	8.6	72	3	Q8WZU1	Q8WZU1 neurospora	929	36.5	8.6	89	15	Q38060	Q38060 human immun
857	36.5	8.6	72	6	Q18889	Q18889 ateles belz	930	36.5	8.6	89	15	Q38061	Q38061 human immun
858	36.5	8.6	72	6	Q18889	Q18889 ateles belz	931	36.5	8.6	89	15	Q38062	Q38062 human immun
859	36.5	8.6	72	11	Q9C2U1	Q9C2U1 mus muscullu	932	36.5	8.6	89	15	Q38065	Q38065 human immun
860	36.5	8.6	72	11	Q9C2U1	Q9C2U1 mus muscullu	933	36.5	8.6	89	15	Q38067	Q38067 human immun
861	36.5	8.6	74	4	Q9CR64	Q9CR64 mus muscullu	934	36.5	8.6	89	15	Q9WA13	Q9WA13 human immun
862	36.5	8.6	75	12	Q41955	Q41955 muirid herpe	935	36.5	8.6	89	16	Q987F9	Q987F9 rhizobium l
863	36.5	8.6	75	16	Q8XL67	Q8XL67 clostridium	936	36.5	8.6	90	16	Q9PFH0	Q9PFH0 xylella fas
864	36.5	8.6	76	12	Q64949	Q64949 avian infec	937	36.5	8.6	90	13	Q9PV59	Q9PV59 acipenser s
865	36.5	8.6	78	2	Q9K1L4	Q9K1L4 streptomyce	938	36	8.5	42	16	Q8X034	Q8X034 ralstonia s
866	36.5	8.6	78	17	Q30275	Q30275 archaeoglob	939	36	8.5	44	2	Q9EX07	Q9EX07 clostridium
867	36.5	8.6	79	15	Q89389	Q89389 human immun	940	36	8.5	47	15	Q87252	Q87252 chimpanzee
868	36.5	8.6	79	15	Q97SH6	Q97SH6 streptococ	941	36	8.5	47	16	Q97L29	Q97L29 clostridium
869	36.5	8.6	79	16	Q98X06	Q98X06 streptomyce	942	36	8.5	48	10	Q9M370	Q9M370 betula ver
870	36.5	8.6	79	17	Q98X06	Q98X06 streptomyce	943	36	8.5	48	10	Q9M370	Q9M370 betula ver
871	36.5	8.6	80	12	Q8QX75	Q8QX75 white spot	944	36	8.5	50	3	Q20195	Q20195 xiphophorus
872	36.5	8.6	80	15	Q90V64	Q90V64 human immun	945	36	8.5	50	3	Q20195	Q20195 caenorhabdi
873	36.5	8.6	80	15	Q90QPL	Q90QPL human immun	946	36	8.5	52	15	Q87372	Q87372 chimpanzee
874	36.5	8.6	80	15	Q905E6	Q905E6 human immun	947	36	8.5	52	16	Q9PGQ3	Q9PGQ3 xylella fas
875	36.5	8.6	81	12	Q36600	Q36600 hepatitis e	948	36	8.5	53	15	Q87347	Q87347 chimpanzee
876	36.5	8.6	81	12	Q36601	Q36601 hepatitis e	949	36	8.5	53	15	Q87402	Q87402 chimpanzee
877	36.5	8.6	81	12	Q36602	Q36602 hepatitis e	950	36	8.5	53	16	Q8U5R8	Q8U5R8 agrobacteri
878	36.5	8.6	81	12	Q36603	Q36603 hepatitis e	951	36	8.5	55	4	Q15281	Q15281 homo sapien
879	36.5	8.6	81	15	Q91893	Q91893 human immun	952	36	8.5	55	12	Q8V0G5	Q8V0G5 asinine her
880	36.5	8.6	81	15	Q930E4	Q930E4 human immun	953	36	8.5	55	12	Q8V0G4	Q8V0G4 asinine her
881	36.5	8.6	81	15	Q98ZV4	Q98ZV4 human immun	954	36	8.5	55	12	Q8V0G3	Q8V0G3 asinine her
882	36.5	8.6	81	15	Q8UPP8	Q8UPP8 human immun	955	36	8.5	55	16	Q9L010	Q9L010 streptomyce
883	36.5	8.6	81	15	Q8UPN8	Q8UPN8 human immun	956	36	8.5	56	4	Q9NVE3	Q9NVE3 homo sapien
884	36.5	8.6	81	15	Q8UPN1	Q8UPN1 human immun	957	36	8.5	56	4	Q96HT7	Q96HT7 homo sapien
885	36.5	8.6	81	15	Q8UPM2	Q8UPM2 human immun	958	36	8.5	56	4	Q95J69	Q95J69 pan troglod
886	36.5	8.6	82	15	Q8UPM0	Q8UPM0 human immun	959	36	8.5	56	12	Q9M9H2	Q9M9H2 equine herp
887	36.5	8.6	82	5	Q8WT04	Q8WT04 plasmodium	960	36	8.5	56	12	Q8V0G2	Q8V0G2 asinine her
888	36.5	8.6	83	11	Q63174	Q63174 rattus norv	961	36	8.5	56	12	Q8V0G1	Q8V0G1 asinine her
889	36.5	8.6	83	15	Q9KX23	Q9KX23 human immun	962	36	8.5	56	12	Q8V0G0	Q8V0G0 asinine her
890	36.5	8.6	83	16	Q99TR3	Q99TR3 staphylococ	963	36	8.5	56	12	Q8V0F9	Q8V0F9 asinine her
891	36.5	8.6	84	9	Q8W6P6	Q8W6P6 bacterioph	964	36	8.5	56	16	Q9ECK2	Q9ECK2 xylella fas
892	36.5	8.6	84	17	Q29898	Q29898 archaeoglob	965	36	8.5	56	16	Q97RW6	Q97RW6 streptococ



```

RA SEQUENCE FROM N.A.
RT xue Q.;
RL "Studies on the tail region of the temperate coliphage 186 genome.";
RN [2] Thesis (1993), University of Adelaide.

RP SEQUENCE FROM N.A.
RX MEDLINE-983711265; PubMed-9705261;
RA Portelli R., Dodd I.B., Xue Q., Egan J.B.;
RT "The late-expressed region of the temperate coliphage 186 genome.";
RL Virology 248:117-130(1998)
DR EMBL; U32222; AAC34169.1;
FT NON_TER 1
FT VARIANT 15 15 S -> *
FT VARIANT 51 51 Q -> *
SQ SEQUENCE 58 AA; 6491 MW; 1199113D8CDEB8E6 CRC64;

Query Match 13.2%; Score 56; DB 9; Length 58;
Best Local Similarity 38.5%; Pred. No. 35;
Matches 10; Conservative 6; Mismatches 10; Indels 0; Gaps 0;

QY 38 PTEALSVAVEEGLAWRKKGCLRLGTH 63
Db 31 PSELYSLSTELITWREKALQSRGNH 56

RESULT 4
ID Q23341 PRELIMINARY; PRT; 80 AA.
AC Q23341;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE Hypothetical 8.5 kDa protein.
GN Z6477.4.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-99069613; PubMed-9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN-BRISTOL N2;
RA None;
RT "The sequence of C. elegans cosmid Z6477.";
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN-BRISTOL N2;
RA Waterston R.;
RT "Direct Submission.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; U40802; AAK19010.1; -.
KW Hypothetical protein.
SQ SEQUENCE 80 AA; 8481 MW; AE43A8268EB6C423 CRC64;

Query Match 12.8%; Score 54; DB 5; Length 80;
Best Local Similarity 28.4%; Pred. No. 89;
Matches 23; Conservative 9; Mismatches 35; Indels 14; Gaps 2;

QY 6 CSSOSISPMBSISENSLVAMDSGOKSRVIENTPEALSVAVEEGLAWRKKGCLRLGTHGS 65
Db 4 CSPKILPAGSSSSSTH-----SSQIRPPLSLASLSEELRVCEGSPRVGAKES 56
QY 66 -----PTASSOSSATNMAI 79
Db 57 SFYCTEQPAQSSYSREDKLCI 77

```

```

RESULT 5
Q9JMT2
ID Q9JMT2 PRELIMINARY; PRT; 84 AA.
AC Q9JMT2;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE YbgA protein.
GN YBG_A.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX Shmizu H., Saitoh Y., Suda Y., Uehara K., Sampei G., Mizobuchi K.;
RT "Complete nucleotide sequence of the F plasmid: Its implications for
organization and diversification of plasmid genomes.";
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE-90317835; PubMed-2164585;
RA Yoshioke Y., Fujita Y., Onosubo E.;
RT "Nucleotide sequence of the promoter-distal region of the tra operon
of plasmid R100, including traI (DNA helicase I) and traD genes.";
RL J. Mol. Biol. 214:39-53(1990).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE-87194554; PubMed-3032897;
RA Saadi S., Maas W.K., Hill D.F., Bergquist P.L.;
RT "Nucleotide sequence analysis of RepFIC, a basic replicon present in
IncFI plasmids P307 and F, and its relation to the RepA replicon of
IncFII plasmids.";
RL J. Bacteriol. 169:1836-1846(1987).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE-95337425; PubMed-7612932;
RA Broom J.E., Hill D.F., Hughes G., Jones W.A., McNaughton J.C.,
Stockwell P.A., Petersen G.B.;
RT "Sequence of a transposon identified as Tn1000 (gamma delta).";
RL DNA Seq. 5:185-189(1995).
RN [5]
RP SEQUENCE FROM N.A.
RA Eichenlaub R.;
RT "F plasmid DNA complete mini-F region (F coordinates 40.301F to
49.869F).";
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE FROM N.A.
RX MEDLINE-86139869; PubMed-3949712;
RA Helsenberg M., Eichenlaub R.;
RT "Twelve 43-base-pair repeats map in a cis-acting region essential for
partition of plasmid mini-F.";
RL J. Bacteriol. 165:1043-1045(1986).
RN [7]
RP SEQUENCE FROM N.A.
RX MEDLINE-99296678; PubMed-10366527;
RA Manwaring N.P., Skurray R.A., Firth N.;
RT "Nucleotide sequence of the F plasmid leading region.";
RL Plasmid 41:219-225(1999).
RN [8]
RP SEQUENCE FROM N.A.
RX MEDLINE-94359430; PubMed-7915817;
RA Frost L.S., Ippen-Ihler K., Skurray R.A.;
RT "An analysis of the sequence and gene products of the transfer region
of the F sex factor.";
RL Microbiol. Rev. 58:162-210(1994).
DR EMBL; AP001918; BAA97888.1; -.
KW Plasmid; Complete proteome.
SQ SEQUENCE 84 AA; 9265 MW; 183C60CAF87121F7 CRC64;

```

```

Query Match          12.58; Score 53; DB 12; Length 80;
Best Local Similarity 25.69; Pred. No. 1.2e+02;
Matches 20; Conservative 12; Mismatches 28; Indels 18; Gaps 64

QY 13 PMRSISNSLVAMDFSGKQSRV-----IENPTALSVAVEEGLAWRKGCCLRLGTHG 64
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Db 6 PVARSGPHSVGELADGKFLGVGRGNLYISEPQARSISSLRGTA-----KHT 55
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

QY 65 SPTASSOSSATNMAIHR 82
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Db 56 SSSSSSSSSQPELIQRS 73
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

RESULT 7
057347 PRELIMINARY; PRT; 88 AA.
ID AC Q57347
AC Q57347;
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Beta B1-crystallin (Fragment);
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97312614; PubMed=9169055;
RA Almann C.R., Chow R.L., Lang R.A., Hemmati-Brivanlou A.;
RT "Lens induction by Pax-6 in Xenopus laevis.";
RL Dev. Biol. 185:119-123(1997).
CC 1- THE VERTEBRATE CRYSTALLINS ARE THE DOMINANT STRUCTURAL COMPONENTS OF
CC 1- THE VERTEBRATE EYE LENS (BY SIMILARITY).
CC 1- SIMILAR GREEK KEY MOTIFS (BY SIMILARITY).
CC 1- SIMILARITY; BELONGS TO THE BETA/GAMMA-CRYSTALLIN FAMILY.
DR EMBL; AF035365; AAB87702.1; -.
DR HSSP; P26775; 1BD7.
DR InterPro; IPR001064; Crystallin.
DR Pfam; PF00030; Crystall; 2.
DR SMART; SM00247; XTALbg. 1.
DR PROSITE; PS00225; CRYSTALLIN.BETAGAMMA; UNKNOWN_1.
KW Duplication; Eye lens protein.
FT NON_TER
FT 1
FT 88
SQ SEQUENCE 88 AA; 10461 MW; A9164E3275EE0E1E CRC64;

Query Match          12.18; Score 51; DB 13; Length 88;
Best Local Similarity 26.69; Pred. No. 2.2e+02;
Matches 17; Conservative 10; Mismatches 23; Indels 14; Gaps 63

QY 3 RSGSGSISPMRSTSENSLVAM---DFSGKQSRVIENTPTALSVAVEEGLAWRKGCCL 63
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Db 14 RSDC-FMSARPIMNOQEHKFLPCTDFKGNKMEIIDDVPSL-----WAYGFCD 63
   1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

QY 59 RLGT 62
   1 1 1
Db 64 RVGS 67
   1 1 1

RESULT 8
09NFC8 PRELIMINARY; PRT; 65 AA.
ID AC Q9NFC8
AC Q9NFC8;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE CG18844 protein.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

```

```

[1]
RP SEQUENCE FROM N.A.
RA Kennedy L.J., Angles J.M.;
RT "Nomenclature for the DLA system, 2000: Second report of the ISAG DLA
RL Nomenclature Committee.";
RD Submitted (JAN-2001) to the EMBL/GenBank/DDJB databases.
DR EMBL; AF343738; AAKI3509.1; -.
FT NON_TER      1
FT              90
SQ SEQUENCE     90 AA;  9600 MW;   D4F5E19F28AD13AF CRC64;

Query Match          11.5%; Score 48.5; DB 7; Length 90;
Best Local Similarity 29.0%; Pred. No. 4.4e+02;
Matches 20; Conservative 8; Mismatches 34; Indels 7; Gaps

QY  2 GRSCGSSQSITPMRISNSLVAMDEFSKGSRVIENTPEALSVAVEGLAHRKKGCLRIG 51
    ||||| | . . . . . ||| ||| | . . . . .
Db   16 GRSGCGSWKDITSITRCSTCASTATWGTSGRSSGGCPT--LSPGTRRRSSWSGR-----G 68
    ||| :
QY  62 THGSPTASS 70
    ||| :
Db   69 PRWTPTADT 77

RESULT 10
QSR6C9 PRELIMINARY; PRT; 89 AA.
AC QSR6C9;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DE TIOR132 protein (Hypothetical protein Atu6153).
GN TIOR132 OR AFU06153 OR AGR_PTI_BX147.
OS Agrobacterium tumefaciens, and
OC Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC Plasmid pTI-SAKURA, and plasmid pTiC58.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=358, 176299;
RN [1]
RP SEQUENCE FROM N.A.
RA SPECIES=A.tumefaciens; STRAIN=MAFF301001; PLASMID=Pti-Sakura;
RX MEDLINE=20184752; PubMed=1072172;
RA Suzuki K., Hattori Y., Uraji M., Ohta N., Iwata K., Murata K.,
RA Katoh A., Yoshida K.;
RT "Complete nucleotide sequence of a plant tumor-inducing Ti plasmid.";
RL Gene 242:331-336(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA SPECIES=A.tumefaciens; STRAIN=MAFF301001; PLASMID=Pti-Sakura;
RX MEDLINE=98193120; PubMed=9524202;
RA Suzuki K., Ohta N., Hattori Y., Uraji M., Katoh A., Yoshida K.;
RT "Novel structural difference between nopaline- and octopine-type trbJ
RT gene: construction of genetic and physical map and sequencing of
RT trbJ/trai and rep gene clusters of a new Ti plasmid pTi-Sakura.";
RL Biochim. Biophys. Acta 1396:1-7(1998).
RN [3]
RP SEQUENCE FROM N.A.
RA SPECIES=A.tumefaciens; STRAIN=MAFF301001; PLASMID=Pti-Sakura;
RX Hattori Y., Suzuki K., Ohta N., Uraji M., Katoh A., Yoshida K.;
RT "Genome structure of pTi-Sakura(I): Strategy for DNA sequencing of a
RT Japanese cherry-Ti plasmid."
RL Nucleic Acids Symp. Ser. 37:159-160(1998).
RN [4]
RP SEQUENCE FROM N.A.
RA SPECIES=A.tumefaciens; STRAIN=MAFF301001; PLASMID=Pti-Sakura;
RX Ohta N., Suzuki K., Hattori Y., Uraji M., Katoh A., Yoshida K.;
RT "Genome structure of pTi-Sakura (III): Characteristics of T-DNA.";
RL Nucleic Acids Symp. Ser. 39:185-186(1998).
RN [5]
RP SEQUENCE FROM N.A.
RA SPECIES=A.tumefaciens; STRAIN=MAFF301001; PLASMID=Pti-Sakura;
RX Uraji M., Suzuki K., Ohta N., Hattori Y., Katoh A., Yoshida K.;
```

```

Query Match          11.5%;   Score 48.5;   DB 5;   Length 65;
Best Local Similarity 37.5%;   Pred. No. 3e+02;
Matches 15;   Conservative 4;   Mismatches 12;   Indels 9;   Gaps 2;

QY 3 RSGCS-----SQSIPMR-SISENSLVAMDFSGQSKR 33
      :::||| | | | | | | | | | | | | | |
Db 12 RRGCTASSKXWALARSVPNSPRESSANFIQVTFQSGPNKR 51

RESULT 9
Q9BD73
ID Q9BD73 PRELIMINARY: PRT; 90 AA.
AC
DT 01-JUN-2001 (TRENDELrel. 17, Created)
DT 01-JUN-2001 (TRENDELrel. 17, Last sequence update)
DT 01-DEC-2001 (TRENDELrel. 19, Last annotation update)
DE MHC class II antigen beta chain (Fragment).
GN DLA-DRB1
OS Canis lupus (Gray wolf).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
NC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis
NCBI TaxID=9612.

```





RA Kalman S., Keating D.H., Palm C., Peck M.C., Surzycki R., Wells D.H.,  
 RA Yeh K.-C., Davis R.W., Federapfel N.A., Long S.R.,  
 RT "Nucleotide sequence and predicted functions of the entire  
 RT Rhodococcus shuttle vectors."  
 RL Rhodococcus shuttle vectors.  
 RL Microbiol. 143:3137-3147(1997).  
 DR EMBL: AF013088; AAC45810.1; -  
 KW Hypothetical protein; plasmid; Complete proteome.  
 SQ SEQUENCE 90 AA; 9914 MW; CC9E8095982DD1D5 CRC64;

Query Match 11.2%; Score 47.5; DB 16; Length 90;  
 Best Local Similarity 22.1%; Pred. No. 5.8e+02;  
 Matches 15; Conservative 15; Mismatches 19; Indels 19; Gaps 3;  
 QY 11 ISPMRSISNSLVAMDFSGQKSRVIENTEALSVAVEEGLAW-----BKKGCLRL-CT 62  
 DB 11 VAPKDLDDSTMSAADFA-----QLFGVYTOGSMSETFYERGQADGIFRLSCT 59  
 QY 63 HGSPTASS 70  
 DB 60 RKTNPSS 67

RESULT 14  
 Q8ROX3  
 ID Q8ROX3 PRELIMINARY; PRT; 79 AA.  
 AC Q8ROX3  
 DT 01-JUN-2002 (TReMBLrel. 21, Created)  
 DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)  
 DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)  
 DE Similar to LOC164714.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=THYMUS GLAND;  
 RA Strausberg R.;  
 RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: BC026208; AHA26208.1;  
 SQ SEQUENCE 79 AA; 8632 MW; 806C30C345C10BE CRC64;

Query Match 11.1%; Score 47; DB 11; Length 79;  
 Best Local Similarity 28.0%; Pred. No. 5.7e+02;  
 Matches 14; Conservative 5; Mismatches 11; Indels 20; Gaps 2;  
 QY 37 NPTALSVAVEEG-----LAWKKKC-----LRLSTHGSP 66  
 DB 16 HPBRLCPSATOGIHAGSLNRRPCTGTLQTFEFSQYHLGERLCTRGAP 65

RESULT 15  
 Q30681  
 ID Q30681 PRELIMINARY; PRT; 86 AA.  
 AC Q30681  
 DT 01-JAN-1998 (TReMBLrel. 05, Created)  
 DT 01-JAN-1998 (TReMBLrel. 05, Last sequence update)  
 DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)  
 DE Plasmid pFAJ2600 putative plasmid multimer resolution  
 DE putative DNA-binding replication protein (repB), putative replication  
 DE protein (repA), and putative plasmid partitioning protein (para)  
 DE genes.  
 OS Rhodococcus erythropolis.  
 OG Plasmid pFAJ2600.  
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
 OC Actinomycetales; Corynebacterineae; Nocardiaceae; Rhodococcus.  
 OX NCBI\_TaxID=1833;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NI86/21;  
 RA MEDLINE=98015402; PubMed=9353918;  
 RA De Mot R., Nagy I., De Schrijver A., Pattanapitpaisal P.,  
 RA Schoofs G., Vanderleyden J.;

RT "Structural analysis of the 6-kb cryptic plasmid pFAJ2600 from  
 RT Rhodococcus erythropolis NI86/21 and construction of Escherichia coli-  
 RT Rhodococcus shuttle vectors."  
 RL Rhodococcus shuttle vectors.  
 RL Microbiol. 143:3137-3147(1997).  
 DR EMBL: AF013088; AAC45810.1; -  
 KW Plasmid.  
 SQ SEQUENCE 86 AA; 9140 MW; CB240A7B86386B18 CRC64;

Query Match 11.1%; Score 47; DB 2; Length 86;  
 Best Local Similarity 22.0%; Pred. No. 6.3e+02;  
 Matches 18; Conservative 18; Mismatches 40; Indels 6; Gaps 3;  
 QY 3 RSCSCSSQISPMRSISNSLVAMDFSGQKSRVIENTEALSVAVEEGLAWKKCGLRG- 61  
 DB 4 RSGTSHPVPRPTGLVAELA-VOMNHPGAGDC--DLTRRMQIWHVHGSSMLRRSGGQAW 60  
 QY 62 --THGSPTASSOSSATMAIHR 81  
 DB 61 RVAHVHTLAEGVNTAAGGVHR 82

RESULT 16  
 Q9RCDA  
 ID Q9RCDA PRELIMINARY; PRT; 79 AA.  
 AC Q9RCDA  
 DT 01-MAY-2000 (TReMBLrel. 13, Created)  
 DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)  
 DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)  
 DE Hypothetical 8.6 kDa protein.  
 OS Xanthomonas campestris.  
 OC Plasmid pKLM43.  
 OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;  
 OC Xanthomonas.  
 OX NCBI\_TaxID=339;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=TAP4-3; TRANSPOSON=TN5044;  
 RX MEDLINE=99406912; PubMed=10476039;  
 RA Minakhina S., Kholodii G., Mindlin S., Yurieva O., Nikiforov V.;  
 RT "TN5053 family transposons are res site hunters sensing plasmid res  
 RT sites occupied by cognate resolvases."  
 RL Mol. Microbiol. 33:1059-1068(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=TAP4-3; TRANSPOSON=TN5044;  
 RA Kholodii G., Yurieva O., Mindlin S., Gorlenko Z., Rybochkin V.,  
 RA Nikiforov V.,  
 RT "TN5044, a novel Tn3 family transposon coding for temperature  
 RT sensitive mercury resistance."  
 RL Res. Microbiol. 151:1-12(2000).  
 DR EMBL: Y17691; CAB65713.1; -  
 KW Hypothetical protein; Plasmid.  
 SQ SEQUENCE 79 AA; 8626 MW; 1639B3E026E36706 CRC64;

Query Match 11.0%; Score 46.5; DB 2; Length 79;  
 Best Local Similarity 31.0%; Pred. No. 6.5e+02;  
 Matches 18; Conservative 14; Mismatches 15; Indels 11; Gaps 4;  
 QY 2 GRSCSSQISPMRSISNSLVAMDFSGQKSRVIB-NPTEA-LSVAVEEGLAWKK 55  
 DB 28 GRKGDLRSFI-----EEAVRAHILELSAQAKVNAHLSEALTDAVDEALAWASK 78

RESULT 17  
 Q96U90  
 ID Q96U90 PRELIMINARY; PRT; 80 AA.  
 AC Q96U90  
 DT 01-DEC-2001 (TReMBLrel. 19, Created)  
 DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)  
 DE 01-MAR-2002 (TReMBLrel. 20, Last annotation update)  
 DE Probable ribosomal protein S19, mitochondrial.  
 GN B109.070.  
 OS Neurospora crassa.

OC Eukaryota; Fungi; ASCOMYCOTA; Pezizomycotina; Sordariomycetes;  
 OC Sordariales; Sordariaceae; Neurospora.  
 RN NCBI\_TaxID=5141;  
 RP SEQUENCE FROM N.A.  
 RA Schulte U., Aign V., Hoheisel J., Brandt P., Fartmann B., Holland R.,  
 RA Nyakatura G., Mewes H.W., Mannhaupt G.;  
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
 [2]  
 RP SEQUENCE FROM N.A.  
 RA German Neurospora genome project;  
 RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AL513409; CAD11378.1; -  
 DR InterPro; IPR002222; Ribosomal\_S19.  
 DR Pfam; PF00203; Ribosomal\_S19; 1.  
 DR ProDom; PD001012; Ribosomal\_S19; 1.  
 KW Ribosomal protein.  
 SQ SEQUENCE 80 AA; 9018 MW; DA38F8D77C20E041 CRC64;  
 Query Match 11.0%; Score 46.5; DB 3; Length 80;  
 Best Local Similarity 25.0%; Pred. No. 6.6e+02;  
 Matches 12; Conservative 14; Mismatches 21; Indels 1; Gaps 1;  
 QY 9 OSISPMRSISENSLVAMDFSGOKSRVIENTPEALSVAVEEGLAWRKKG 56  
 DB 30 KKIAPIRQARSAILPNFVGLKQV-ENGKDYIDLTVTEEMVGHKLK 76  
 RESULT 18  
 Q98KZ6 PRELIMINARY; PRT; 88 AA.  
 AC Q98KZ6;  
 DT 01-OCT-2001 (TrEMBLrel. 18, Created)  
 DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
 DE Hypothetical protein msr1249.  
 GN MSR1249.  
 OS Rhizobium loti (Mesorhizobium loti).  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 OC Phyllobacteriaceae; Mesorhizobium.  
 OX NCBI\_TaxID=381;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MAFF303099;  
 RX MEDLINE=21082930; PubMed=11214968;  
 RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,  
 RA Watanabe A., Ikesawa K., Ishikawa A., Kawashima K., Kimura T.,  
 RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,  
 RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpō S., Sugimoto M.,  
 RA Takeuchi C., Yamada M., Tabata S.;  
 RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium  
 Mesorhizobium loti.";  
 RL DNA Res. 7:331-338(2000).  
 DR EMBL; AP002997; BAB48667.1; -  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 88 AA; 9552 MW; DA4EE5FE549D2D2C CRC64;  
 Query Match 11.0%; Score 46.5; DB 16; Length 88;  
 Best Local Similarity 22.0%; Pred. No. 7.4e+02;  
 Matches 11; Conservative 13; Mismatches 25; Indels 1; Gaps 1;  
 QY 14 MRSISENSLVAMDFSGOKSRV-IENTPEALSVAVEEGLAWRKKGLRGT 62  
 DB 28 WSLLSLGRIATEFSARARYOTERAIRSLPIELQKDIGWPEASDFTKGT 77  
 RESULT 19  
 Q93JW8 PRELIMINARY; PRT; 90 AA.  
 AC Q93JW8;  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)

DE Cysteine proteinase (Fragment).  
 OS Narcissus pseudonarcissus (Daifodil).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Asparagales; Amaryllidaceae;  
 OC Narcissus.  
 OX NCBI\_TaxID=39639;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. DUTCH MASTER; TISSUE=SENESCING TEPALS OF 4-DAY OLD FLOWERS;  
 RA Hunter D.A., Yi M., Reid M.S.;  
 RT "Role of Ethylene and ABA in Perianth Senescence of Daffodil  
 (Narcissus pseudonarcissus L. Dutch Master).";  
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF420011; AAL16903.1; -  
 DR InterPro; IPR002086; Aldehyde\_dehydr.  
 DR InterPro; IPR000668; Peptidase\_C1.  
 DR Pfam; PF00112; Peptidase\_C1; 1.  
 DR ProDom; PD000158; Peptidase\_C1; 1.  
 DR PROSITE; PS00687; ALDEHYDE\_DEHYDR\_GLU; UNKNOWN\_1.  
 FT NON\_TER 1 90  
 FT NON\_TER 90 90  
 SQ SEQUENCE 90 AA; 9346 MW; 5B7C407359B3A658 CRC64;  
 Query Match 11.0%; Score 46.5; DB 10; Length 90;  
 Best Local Similarity 33.3%; Pred. No. 7.6e+02;  
 Matches 24; Conservative 13; Mismatches 22; Indels 13; Gaps 7;  
 QY 1 QGRSGCSQSISPMRSISENSLVAMDFSGOKSRVIENTPEALSVAVEEGLAWR--KKG 57  
 DB 19 OGR--CSSKK-SPIVSIDGYQNVNNEAALMKAVAN--QPVSAIEASGMAFQYSEGV 73  
 QY 58 L--RLGT---HG 64  
 DB 74 FTGRCGTLDHG 85  
 RESULT 20  
 Q8VY75 PRELIMINARY; PRT; 54 AA.  
 AC Q8VY75;  
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
 DE Putative MYB family transcription factor.  
 GN AT2G16720.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Nguyen M., Karlin-Neumann G., Southwick A., Lam B., Miranda M.,  
 RA Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,  
 RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,  
 RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,  
 RA Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,  
 RA Ecker J., Theologis A., Davis R.W.;  
 RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AY072387; AAL62379.1; -  
 SQ SEQUENCE 54 AA; 6078 MW; 2A9E672F34B7BEDA CRC64;  
 Query Match 10.9%; Score 46; DB 10; Length 54;  
 Best Local Similarity 31.7%; Pred. No. 4.6e+02;  
 Matches 13; Conservative 9; Mismatches 9; Indels 10; Gaps 2;  
 QY 6 CSSQSI----SPMRSENSL-----VAMDFSGOKSRVIE 36  
 DB 7 CSSETVKCQTENSSSISSIDISSNVGYDFLGLKTRILD 47  
 RESULT 21  
 Q76048

```

ID Q76048 PRELIMINARY; PRT; 77 AA.
AC Q76048;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-NAR-2002 (TReMBLrel. 20, Last annotation update)
DE P24/p25/p7 (Fragment).
GN GAG.
OS Human immunodeficiency virus type 1.
OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RA Salmunen M.;
RP SEQUENCE FROM N.A.
RT "Rapid and simple characterization of in vivo HIV-1 sequences using solid-phase direct sequencing.";
RL AIDS Res. Hum. Retroviruses 8:1733-1742(1992).
[2]
RN HSSP: P05888; IAAF.
DR InterPro: IPR001878; Znf.CCHC.
DR Pfam: PF00098; Zf.CCHC; 2.
DR PRINTS: PR00939; C2HCZNFINGER.
DR SMART: SM00343; Znf.C2HC; 2.
FT NON_TER 1
FT NON_TER 77
SQ SEQUENCE 77 AA; 8724 MW; 2BEAEE9B5C3E6A6 CRC64;

Query Match 10.9%; Score 46; DB 15; Length 77;
Best Local Similarity 10.9%; Pred. No. 7.2e+02;
Matches 15; Conservative 7; Mismatches 14; Indels 28; Gaps 2;

QY 29 GQSRVTVT-----EPTALSVAVEEG-----LAWKKGCLRL 60
DB 1 GHKARVLAQMSKATNAATIMRGFRNORKTVKFCNGKGGHIAIARNCRAPRKGGWC 60
QY 61 GTGH 64
DB 61 GKEG 64

RESULT 22
Q8WME9 PRELIMINARY; PRT; 86 AA.
AC Q8WME9;
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
DE Peptidoglycan recognition protein (Fragment).
GN PGLYRP.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RA TISSUE=ADIPOSE TISSUE;
RP SEQUENCE FROM N.A.
RT Kappeler S.R.;
RT "The peptidoglycan recognition protein, PGRP, is expressed in the lactating mammary gland of camels.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ310355; CAC83647.1;
FT NON_TER 1
FT CHAIN 1
FT NON_TER 86
FT NON_TER 86
SQ SEQUENCE 86 AA; 9517 MW; BFC5DC02AF11CBF5 CRC64;

Query Match 10.9%; Score 46; DB 6; Length 86;

```

```

Best Local Similarity 38.1%; Pred. No. 8.2e+02;
Matches 8; Conservative 4; Mismatches 9; Indels 0; Gaps 0;

QY 47 EEWLRKKKGCLRLGTHGSPT 67
DB 39 EDGLVEGGRGNTVGAHSGPT 59

RESULT 23
Q9PHK1 PRELIMINARY; PRT; 87 AA.
AC Q9PHK1;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-NAR-2002 (TReMBLrel. 20, Last annotation update)
DE Hypothetical protein Xfa0004.
GN XFA0004
OS Xylella fastidiosa.
OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
OC Xylella.
OX NCBI_TaxID=2371;
RN [1]
RA SEQUENCE FROM N.A.
RC STRAIN=9A5C;
RX MEDLINE=20365717; PubMed=10910347;
RA Simpson A.J.G., Reinach F.C., Artuda P., Abreu F.A., Acencio M.,
RA Alvarenga R., Alves L.M.C., Araya J.E., Bala G.S., Baptista C.S.,
RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carriao D.M., Carrier H.,
RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,
RA Facinani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA Foga J.S., Franca S.C., Franco M.C., Frohme M., Gomes S.L., Gruber A.,
RA Gargier M., Goldman G.H., Goldman M.H.S., Kemper E.L., Kitzajima J.P.,
RA Ho P.L., Honeisel J.D., Junqueira M.L., Kempner E.L., Lambais M.R., Leite L.C.C.,
RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Lopes L.C.C.,
RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
RA Marquês M.V., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA Nhani A.Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA Peixoto B.R., Pereira G.A.G., Pereira H.A.Jr., Pesquero J.B.,
RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsubako M.H.,
RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
RT "The genome sequence of the plant pathogen Xylella fastidiosa.";
RL Nature 406:151-159(2000).
DR EMBL: AF003851; AAF85573.1;
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 87 AA; 10116 MW; F9391C1497D962FF CRC64;

Query Match 10.9%; Score 46; DB 16; Length 87;
Best Local Similarity 28.1%; Pred. No. 8.3e+02;
Matches 16; Conservative 10; Mismatches 23; Indels 8; Gaps 3;

QY 14 MRSISENSIVAMDF----SGQSRVIENTPEALSVAVEEGKGLRGLTGSP 66
DB 1 MNTISEELA--DFVKQAVRKLHIQNDAEKYEIFAT--LTWKEGDHLYVTRGKP 53

RESULT 24
Q9ZXL4 PRELIMINARY; PRT; 89 AA.
AC Q9ZXL4
DT 01-MAY-1999 (TReMBLrel. 10, Created)
DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)

```

[illegible]

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 28, 2003, 08:56:00 ; Search time 18.5197 Seconds  
(without alignments)  
188.125 Million cell updates/sec

Sequence: 1 QGRSCSSQSIIPNRISSEN.....SPTASSQSSATNMAIHRSP 84  
P10132 MYCOPASMA

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 7863

Minimum DB seq length: 40  
Maximum DB seq length: 90

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 1000 summaries

Database: SwissProt 40

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	51.5	12.2	73	1	RPON_METJA
2	47	11.1	71	1	Y16K_BPT4
3	46.5	11.0	62	1	Y203_METJA
4	45	10.6	76	1	CD24_RAT
5	45	10.6	79	1	MT2_MALDO
6	45	10.6	82	1	RADC_STAAD
7	44.5	10.5	68	1	GNGL_HUMAN
8	43	10.2	82	1	Y567_METJA
9	42.5	10.0	67	1	HFO2_METFO
10	42.5	10.0	88	1	RS19_CHLTR
11	42	9.9	45	1	AT12_HSVE4
12	42	9.9	78	1	YHGG_ECOLI
13	42	9.8	83	1	TMOB_PSEME
14	41.5	9.8	67	1	HFOB_METFO
15	41.5	9.8	87	1	RS19_MYCGA
16	41.5	9.8	88	1	C552_MARHY
17	41.5	9.8	88	1	RS19_CHLNP
18	41	9.7	65	1	RPON_SULSO
19	40.5	9.6	43	1	CC3_CARCN
20	40.5	9.6	68	1	GB05_HUMAN
21	40	9.5	41	1	HS2K_LICES
22	40	9.5	55	1	RPON_THEAC
23	40	9.5	72	1	Y187_BPT7
24	40	9.3	83	1	CC4_CARCN
25	39.5	9.3	43	1	RS22_ECOLI
26	39	9.2	45	1	CC4_CARCN
27	39	9.2	43	1	RS22_ECOLI
28	39	9.2	85	1	R37A_MYXGL
29	38.5	9.1	54	1	BAF_BORBR
30	38.5	9.1	67	1	CSPT_STRCO
31	38.5	9.1	72	1	Y003_ARCFU
32	38.5	9.1	74	1	COSA_BOVIN
33	38.5	9.1	74	1	NIFH_NOSSN

34	38.5	9.1	78	1	CINA_STRGV
35	38.5	9.1	82	1	S6B1_FEAST
36	38.5	9.1	87	1	RS19_MYCPN
37	38.5	9.1	88	1	RS19_MYCCA
38	38.5	9.1	88	1	SMFX_HUMAN
39	38	9.0	54	1	IOVO_DRONO
40	38	9.0	58	1	NINF_BPP22
41	38	9.0	59	1	SINI_BACLI
42	38	9.0	59	1	FER_METBA
43	38	9.0	63	1	FER2_DESYM
44	38	9.0	67	1	YDFZ_ECOLI
45	38	9.0	88	1	Y874_CAUCR
46	37.5	8.9	88	1	RS19_MYCGE
47	37.5	8.9	88	1	DADR_MOUSE
48	37.5	8.9	88	1	RS19_UREPA
49	37	8.7	53	1	LECA_LATAP
50	37	8.7	54	1	IOVO_CASCA
51	37	8.7	60	1	NXSL_DENVI
52	37	8.7	66	1	CSP7_STRCL
53	37	8.7	67	1	HMT1_METH
54	37	8.7	76	1	TPKG_MOUSE
55	37	8.7	88	1	V685_BPHLS
56	37	8.7	89	1	ALB1_FHANG
57	36.5	8.6	80	1	Y509_ECO57
58	36.5	8.6	84	1	FER_BUTWE
59	36	8.5	55	1	YH13_ARCFU
60	36	8.5	59	1	NXSL_DENJA
61	36	8.5	60	1	COXO_HUMAN
62	36	8.5	63	1	SR14_MACRA
63	36	8.5	74	1	IPKG_HUMAN
64	36	8.5	76	1	GCH1_MUCHA
65	36	8.5	80	1	RS16_VIBCH
66	36	8.5	82	1	TRBG_ECOLI
67	36	8.5	84	1	SCX2_CENNO
68	36	8.5	85	1	NEUL_PAPHA
69	36	8.5	87	1	SCX3_CENSC
70	36	8.5	87	1	SCX3_CENSC
71	36	8.5	88	1	Y408_RHISN
72	36	8.5	88	1	Y408_RHISN
73	35.5	8.4	61	1	BD02_RAT
74	35.5	8.4	63	1	GBGU_BOVIN
75	35.5	8.4	69	1	MT1_CASGL
76	35.5	8.4	71	1	ATP9_PARTE
77	35.5	8.4	75	1	TM08_ARATH
78	35.5	8.4	77	1	PSAC_MASLA
79	35.5	8.4	80	1	TM08_YEAST
80	35.5	8.4	87	1	RL31_SULSO
81	35.5	8.4	88	1	VIT_MELGA
82	35	8.3	50	1	RL40_AERPE
83	35	8.3	52	1	RL40_LEIMA
84	35	8.3	52	1	CK5A_CONPU
85	35	8.3	67	1	HMT2_METH
86	35	8.3	67	1	ICIC_HIRME
87	35	8.3	70	1	RL29_TREPA
88	35	8.3	73	1	CAT2_BOVIN
89	35	8.3	75	1	ME10_EUPRA
90	35	8.3	76	1	BB11_SCHCO
91	35	8.3	77	1	YCXB_CVAPA
92	35	8.3	79	1	NSGX_HUMAN
93	35	8.3	82	1	CUD5_SCHGR
94	35	8.3	90	1	YEAC_ECOLI
95	35	8.3	90	1	MUTI_ENTMU
96	34.5	8.2	43	1	RPON_METTH
97	34.5	8.2	55	1	BD03_MOUSE
98	34.5	8.2	63	1	ACAL_ACALI
99	34.5	8.2	71	1	YHDL_ECOLI
100	34.5	8.2	72	1	PL24_HELPY
101	34.5	8.2	73	1	MT1_COPAR
102	34.5	8.2	80	1	YHDL_HVIMA
103	34.5	8.2	82	1	YHDL_HVIMA
104	34.5	8.2	82	1	YHDL_HVIMA
105	34.5	8.2	82	1	YHDL_HVIMA
106	34.5	8.2	88	1	PTHP_BACST

P29827	streptovirt
P52870	saccharomyc
P75576	mycoplasma
P10132	mycoplasma
Q9uhp9	homo sapien
P05560	dromalus no
Q38666	bacterioph
P22755	bacillus li
P00202	methanosarc
P10624	desulfovibr
P76153	eschericich
P58091	caulobacter
P47401	mycoplasma
Q61616	mus musculu
Q9pqg6	ureaplasma
P07441	lathyrus ap
P05559	casuarina c
P01418	dendroaspis
Q01761	streptomyce
P50483	methanobact
Q70139	mus musculu
Q03302	mycobacteri
Q9ftr8	phaseolus a
P58092	eschericich
P12732	haloarcu
P14073	butyrilact
Q28560	archaeoglob
P01417	dendroaspis
P15954	homo sapien
O18881	macaca radi
Q9Y2B9	homo sapien
P51598	mucuna hass
Q9ku90	vibrio chol
P41072	eschericich
P01495	centruroid
P32005	papio hamad
P01493	centruroid
P01494	centruroid
P55597	rhizobium s
O30153	archaeoglob
Q88514	rattus norv
P50154	bos taurus
Q39511	casuarina g
P66001	parametium
Q29974	arabidopsis
O07112	mastigoclad
P37744	saccharomyc
P58190	sulfolobus
P36531	meleagris g
Q9yfy7	aeropyrum p
Q05551	leishmania
Q9u626	conus purpu
Q27731	methanobact
P01051	hirudo medi
O83227	treponema p
P05689	bos taurus
P12350	euplotes ra
P78742	schizophril
P48332	cyanoophora
Q9uh64	homo sapien
P56562	schistocerc
P76231	eschericich
P80925	enterococc
Q26147	methanobact
Q9wt10	mus musculu
P81592	acalolepta
P36675	eschericich
P56045	helicobacte
P43396	coffea arab
P05924	human immun
P16420	anabaena cy
P41192	serratia ma
P42013	bacillus st

107	34.5	8.2	89	1	RS14_STRPN	Q97sv0 streptococc	180	32.5	7.7	78	1	GVPA_HALME	P23761 halobacteri
108	34.5	8.2	89	1	RS19_XLFA	Q9pe72 xylella fas	181	32.5	7.7	78	1	GVPA_NATVU	O33397 natronobact
109	34.5	8.2	90	1	RS19_BRAJA	P10335 bradyrhizob	182	32.5	7.7	79	1	MT2_ACTCH	O43390 actinidia c
110	34	8.0	46	1	V11_BPT3	P20833 bacterioph	183	32.5	7.7	79	1	AFB3_RAPSA	O24332 raphanus sa
111	34	8.0	50	1	VLN_ARPEE	Q9v9t9 aeropyrum p	184	32.5	7.7	80	1	V906_TREPA	O83876 treponema p
112	34	8.0	51	1	VG62_BEMD2	O64254 mycobacteri	185	32.5	7.7	81	1	VP0_HV1JR	P20882 human immun
113	34	8.0	52	1	RL40_TRYBB	P12899 trypanosoma	186	32.5	7.7	81	1	VP0_HV1SI	P19554 human immun
114	34	8.0	52	1	RL40_TRYCR	P14795 trypanosoma	187	32.5	7.7	83	1	C554_PARSF	P00105 paracoccus
115	34	8.0	52	1	Y07K_FRY	P20954 papaya mosa	188	32.5	7.7	83	1	Y07S_BRUME	Q8y717 bruceella me
116	34	8.0	71	1	EXY5_STRPY	Q93774 streptococc	189	32.5	7.7	87	1	Y285_ARCFU	O29984 archaeoglob
117	34	8.0	74	1	NXLA_BUNMU	P01378 bungalow mu	190	32.5	7.7	87	1	Y898_MYCTU	O10566 mycobacteri
118	34	8.0	74	1	Y8K2_SSV1	P20394 vaccinia vi	191	32.5	7.7	87	1	YFJM_ECOLI	P52128 escherichia
119	34	8.0	78	1	YH65_SYNY3	P72913 synecocyst	192	32.5	7.7	87	1	HG17_CHICK	P02314 gallus gall
120	34	8.0	79	1	VPW_OMVVS	P19510 ovine lentil	193	32.5	7.7	89	1	RS20_HELPJ	O92mz1 helicobacte
121	34	8.0	86	1	Y165_SYNP6	P81290 bacillus st	194	32.5	7.7	89	1	RS20_HELPV	P56027 helicobacte
122	34	8.0	88	1	RS16_BACST	P33801 bacterioph	195	32.5	7.7	90	1	PTSO_ECOLI	P33996 escherichia
123	34	8.0	88	1	Y67_BPT7	P33084 bacterioph	196	32.5	7.7	90	1	RHT3_RHOTO	P20466 rhodospirid
124	34	8.0	89	1	Y12E_BPT4	Q92ct1 rickettsia	197	32	7.6	47	1	YK82_ARCFU	O28197 archaeoglob
125	34	8.0	90	1	RS2D_RICPR	O48425 bacterioph	198	32	7.6	52	1	SRV_AKOAZ	P36388 akodon azar
126	33.5	7.9	58	1	N1NF_BPH19	Q95705 hylobates l	199	32	7.6	54	1	Y0VQ_RHEAM	P05558 rhea americ
127	33.5	7.9	68	1	ATP8_HYLLA	P01522 conus geogr	200	32	7.6	57	1	Y058_NPVAC	P41462 autographa
128	33.5	7.9	73	1	CX06_CONGE	P05676 synecococc	201	32	7.6	58	1	RS21_CLOPE	O8xiu0 clostridium
129	33.5	7.9	77	1	Y165_SYNP6	O96804 skeletonema	202	32	7.6	60	1	NXSL_ASTST	P01438 astrotia st
130	33.5	7.9	80	1	PSAC_SRECO	P50706 mus musculu	203	32	7.6	62	1	V452_BPT4	P07878 bacterioph
131	33.5	7.9	81	1	DEF8_MOUSE	P49477 odontella s	204	32	7.6	63	1	YDIE_ECOLI	P40721 escherichia
132	33.5	7.9	81	1	PSAC_ODOSI	O64016 mus musculu	205	32	7.6	64	1	PHYB_PHYBI	P81565 phylomedusa
133	33.5	7.9	82	1	DEPH_MOUSE	P56509 methanobact	206	32	7.6	64	1	YEWE_ECOLI	P76366 escherichia
134	33.5	7.9	82	1	YFTD_METFE	P50893 platymonas	207	32	7.6	65	1	RK33_MAIZE	P25461 zea mays (m
135	33.5	7.9	83	1	RT19_PLASU	O00479 homo sapien	208	32	7.6	66	1	CADO_BUNCA	P81783 bungalow ca
136	33.5	7.9	83	1	NER_HUMAN	P50893 platymonas	209	32	7.6	66	1	HST_YERKR	P31518 yersinia kr
137	33.5	7.9	89	1	V45_BPT7	O00479 homo sapien	210	32	7.6	70	1	GBG1_DROME	P38040 drosophila
138	33.5	7.9	89	1	CX3_CONGE	P46496 haemophilus	211	32	7.6	70	1	YVAK_VACCC	P20520 vaccinia vi
139	33	7.8	42	1	CY06_PROHO	P33785 bacterioph	212	32	7.6	71	1	ICIA_SOLTU	P01052 solanum tub
140	33	7.8	45	1	LHA_HOMA	P58924 conus geogr	213	32	7.6	71	1	YQGC_ECOLI	P46878 escherichia
141	33	7.8	52	1	JUJA_ENTAE	P81244 prochloroth	214	32	7.6	73	1	VF14_VAVV	P33872 variola vir
142	33	7.8	56	1	Q939_BSPFI	P80259 rhodospirid	215	32	7.6	75	1	MCHB_ECOLI	Q9rm53 escherichia
143	33	7.8	61	1	CSRA_PSEAE	O68413 bacterioph	216	32	7.6	75	1	RPOK_ARCFU	O29134 archaeoglob
144	33	7.8	61	1	CSRA_PSEFL	O69078 pseudomonas	217	32	7.6	75	1	RPOZ_PORPU	O29136 porphyra pu
145	33	7.8	62	1	GBG_CAEEL	Q9x606 caenorhabdi	218	32	7.6	76	1	RPON_ARCFU	O29135 archaeoglob
146	33	7.8	62	1	GBG_CAEEL	P54406 caenorhabdi	219	32	7.6	76	1	TXP4_APTSC	P49269 aptostichus
147	33	7.8	72	1	GBGE_CALVI	Q9nt22 calliphora	220	32	7.6	77	1	RHT2_RHOTO	P20465 rhodospirid
148	33	7.8	72	1	GBGE_DROME	Q9nt23 drosophila	221	32	7.6	80	1	CD24_HUMAN	P25063 homo sapien
149	33	7.8	72	1	HAF1_ARCFU	O29910 archaeoglob	222	32	7.6	80	1	SAP_PIG	P81405 sus scrofa
150	33	7.8	72	1	HSTC_YEREN	O50319 yersinia en	223	32	7.6	81	1	RS16_CLOAB	Q97197 clostridium
151	33	7.8	73	1	IAA1_STROI	P09921 streptomyce	224	32	7.6	81	1	YAHM_ECOLI	P75692 escherichia
152	33	7.8	75	1	IAA2_STROI	P20596 streptomyce	225	32	7.6	82	1	CUD5_LOCOLI	P56561 locusta mig
153	33	7.8	75	1	YODD_ECOLI	P76328 escherichia	226	32	7.6	82	1	YCIN_ECOLI	P46132 locusta mig
154	33	7.8	78	1	GLPE_HUMAN	P15421 homo sapien	227	32	7.6	83	1	YCX1_PINTH	Q00865 pinus thunb
155	33	7.8	78	1	RPOH_HUMAN	Q28915 pan troglod	228	32	7.6	84	1	RS21_HELPHI	P38654 entamoeba h
156	33	7.8	78	1	RUXX_METJA	Q58443 methanobact	229	32	7.6	86	1	TRLL_HELPHI	O87326 helicobacte
157	33	7.8	81	1	RUXX_METTH	O26745 methanobact	230	32	7.6	86	1	EXE4_HELPSU	P26349 helicobacte
158	33	7.8	81	1	TRUA_SYNP6	O24712 synecococc	231	32	7.6	87	1	NXLB_BUNMU	P15818 bungalow mu
159	33	7.8	82	1	RS16_PASMU	Q58123 pasteurella	232	32	7.6	87	1	SCX3_CENSC	Q95wd2 centruloide
160	33	7.8	82	1	YK3_CAEEL	P819x1 thermotanaer	233	32	7.6	87	1	SCX4_CENNO	P5682 centruloide
161	33	7.8	82	1	YK3_CAEEL	Q03561 caenorhabdi	234	32	7.6	87	1	SCX5_CENNO	P40129 erwinia chr
162	33	7.8	83	1	YCY2_PARDE	P29969 paracoccus	235	32	7.6	89	1	HEM3_ERWCH	O83053 treponema p
163	33	7.8	84	1	YF7A_PSEAE	P58040 pseudomonas	236	32	7.6	89	1	Y008_TREPA	O83053 treponema p
164	33	7.8	85	1	MTRE_METVA	Q50830 methanobact	237	32	7.6	89	1	YHHL_ECOLI	P37614 escherichia
165	33	7.8	85	1	NAPD_ECOLI	P33938 escherichia	238	32	7.6	90	1	RS20_FUSNN	Q8mhw1 fusobacteri
166	33	7.8	89	1	VPR_STVSA	P12521 simian immu	239	32	7.6	90	1	N1NF_BPP21	Q9xjq3 bacterioph
167	33	7.8	90	1	RS16_RICRI	O68451 rickettsia	240	31.5	7.4	57	1	V3A_IBVM5	P02137 avian infec
168	32.5	7.7	45	1	VPUL_IV123	Q9kall bacillus ha	241	31.5	7.4	57	1	V3A_IBVM5	P02137 avian infec
169	32.5	7.7	47	1	RECA_PSEST	P08805 human immun	242	31.5	7.4	60	1	PEPA_URSTH	P13636 ursus thibe
170	32.5	7.7	52	1	RL33_CANJE	Q07809 pseudomonas	243	31.5	7.4	62	1	RS27_METAC	Q8tsz8 methanosarc
171	32.5	7.7	52	1	MT_SYNP7	Q9p138 campylobact	244	31.5	7.4	68	1	RL29_PYRAB	Q9vlu2 pyrococcus
172	32.5	7.7	55	1	V3A_IBVM4	P17041 homo sapien	245	31.5	7.4	68	1	RL29_PYRAB	Q9vlu2 pyrococcus
173	32.5	7.7	55	1	V3A_IBVM4	P30331 synecococc	246	31.5	7.4	70	1	YATM_RHISN	P5513 rhizobium s
174	32.5	7.7	58	1	CKKN_BOVIN	P30259 avian infec	247	31.5	7.4	72	1	CCAZ_ECOLI	O46995 escherichia
175	32.5	7.7	58	1	ATPE_YEAST	P41520 bos taurus	248	31.5	7.4	72	1	YQ65_LISIN	Q8u005 pyrococcus
176	32.5	7.7	61	1	HF01_METFO	P21306 saccharomyc	249	31.5	7.4	74	1	YQ65_LISIN	Q926b0 listeria in
177	32.5	7.7	67	1	YLDH_ALCEU	P48782 methanobact	250	31.5	7.4	75	1	PSK3_ORYSA	Q9frf9 oryza sativ
178	32.5	7.7	77	1	CX04_CONST	P40289 alcaligenes	251	31.5	7.4	79	1	Y124_AQOEA	O65524 aquifex aeo
179	32.5	7.7	78	1	CX04_CONST	Q9xzk3 conus stria	252	31.5	7.4	80	1	APF2_RAPSA	P30230 raphanus sa

253	31.5	7.4	80	1	FRL1_METJA	Q60368 methanococc	326	30.5	7.2	47	1	RL33_STAAB	Q9ags7 staphylococ
254	31.5	7.4	80	1	PSAC_ANTSP	Q06439 antithamnio	327	30.5	7.2	54	1	PM01_LOCM1	P80059 locusta miq
255	31.5	7.4	80	1	HWAC_SCHPO	Q13916 schizosacch	328	30.5	7.2	57	1	V3A_IBVP3	P30238 avian infec
256	31.5	7.4	81	1	ULIL_HSV62	P52459 human herpe	329	30.5	7.2	58	1	FER_METTE	Q01700 methanosarc
257	31.5	7.4	81	1	YLQC_BACSU	Q31738 bacillus su	330	30.5	7.2	59	1	PART_HUMAN	Q9npd0 homo sapien
258	31.5	7.4	82	1	MT2B_LYCES	Q40158 lycopersico	331	30.5	7.2	60	1	NXSL_MICNI	P80548 micrurus ni
259	31.5	7.4	82	1	VPU_HVIBN	P20212 human immu	332	30.5	7.2	60	1	Y05L_BPT4	P39237 bacterioph
260	31.5	7.4	84	1	Y101_SSV1	P20212 sulfolobus	333	30.5	7.2	62	1	RHT1_RHOTO	P07907 rhodospirid
261	31.5	7.4	86	1	Y009_BPHP1	P51710 bacterioph	334	30.5	7.2	63	1	CSRA_HAELN	P44879 haemophilus
262	31.5	7.4	87	1	DEFA_RAT	Q62713 rattus norv	335	30.5	7.2	63	1	RP09_FOHPV	Q9J5a4 fowipox vir
263	31.5	7.4	87	1	DEFB_RAT	Q921f1 rattus norv	336	30.5	7.2	63	1	RS21_BURMA	Q9w25 burkholderi
264	31.5	7.4	87	1	IM13_ARATH	Q9xh48 arabidopsis	337	30.5	7.2	69	1	NP08_MACRO	P92663 macropus ro
265	31.5	7.4	88	1	RS20_CLOAB	Q971k0 clostridium	338	30.5	7.2	70	1	FHFL_AERHI	P19369 aeromonas h
266	31.5	7.4	89	1	HG17_HUMAN	P05204 homo sapien	339	30.5	7.2	70	1	REG1_BOVIN	P82943 bos taurus
267	31.5	7.4	89	1	HLA_BACSU	P39798 bacillus su	340	30.5	7.2	70	1	RS21_BURPS	P70943 burkholderi
268	31.5	7.4	90	1	REV_HV1W2	P05866 human immu	341	30.5	7.2	71	1	Y029_NPVAC	P41433 autographa
269	31	7.3	43	1	CC2_CARCN	P32955 carica cand	342	30.5	7.2	71	1	CB0Q_CHRVI	P65440 chromatium
270	31	7.3	46	1	VIT_ANAPL	P56530 anas platyr	343	30.5	7.2	74	1	CSMB_CHLVI	Q46473 chlorobium
271	31	7.3	48	1	RL40_METTH	Q26653 methanobact	344	30.5	7.2	75	1	GVAL_HALNI	P08958 halobacteri
272	31	7.3	54	1	GALA_CANFA	P33710 catfish faml	345	30.5	7.2	77	1	TACL_TACTR	P14213 tachypleus
273	31	7.3	55	1	LECA_LATSA	P23208 lathyrus sa	346	30.5	7.2	77	1	YOR2_FIVSD	P19033 feline immu
274	31	7.3	56	1	VG36_BMD2	Q22001 mycobacteri	347	30.5	7.2	78	1	CBX2_HUMAN	Q14781 homo sapien
275	31	7.3	60	1	TAP_ORNMO	P47726 ornithodor	348	30.5	7.2	80	1	PSAC_CHLRE	Q00914 chlamydomon
276	31	7.3	62	1	UCRX_BOVIN	P00130 bos taurus	349	30.5	7.2	80	1	PSAC_CVACA	Q19905 cyanidium c
277	31	7.3	63	1	JHEB_TRINI	P30810 trichoplusi	350	30.5	7.2	80	1	PSAC_PINTH	P19649 pinus thum
278	31	7.3	63	1	YKIA_ECOLI	P08366 escherichia	351	30.5	7.2	81	1	SPR6_RABIT	P12517 human immu
279	31	7.3	63	1	RA35_CVACA	Q9t1r9 cyanidium c	352	30.5	7.2	81	1	R27A_YEAST	P18776 oryctolagus
280	31	7.3	66	1	COPP_HELPJ	Q92m70 helicobacte	353	30.5	7.2	82	1	R27B_YEAST	P35907 saccharomyc
281	31	7.3	66	1	DMS5_PACD	Q93455 pectinmedusa	354	30.5	7.2	82	1	RL27_SPRO	Q11111 saccharomyc
282	31	7.3	66	1	RPON_SULAC	P39472 sulfolobus	355	30.5	7.2	84	1	RS20_BORBU	Q95394 borrelia bu
283	31	7.3	67	1	ATP8_MOUSE	P03930 mus musculu	356	30.5	7.2	85	1	SCXA_CENSC	P26287 centruoide
284	31	7.3	67	1	IF1A_MOUSE	Q60872 mus musculu	357	30.5	7.2	85	1	PER_DROTE	Q26287 centruoide
285	31	7.3	68	1	YHBC_KLEPN	P11160 klebsiella	358	30.5	7.2	86	1	RS12_MYCGO	P20000 mycobacteri
286	31	7.3	69	1	RL28_AQUAE	Q66492 aquifex aeo	359	30.5	7.2	88	1	RS20_AGRIT5	Q8u1h2 agrobacteri
287	31	7.3	70	1	RL31_HAEDU	Q59450 haemophilus	360	30.5	7.2	88	1	RS20_AGRIT5	P02313 bos taurus
288	31	7.3	72	1	Y541_BORBU	Q51491 borrelia bu	361	30.5	7.2	89	1	HG17_BOVIN	P80272 sus scrofa
289	31	7.3	73	1	BP4C_BRANA	P41506 brassica na	362	30.5	7.2	89	1	HG17_PIG	Q27304 methanobact
290	31	7.3	73	1	YB76_HAELN	P44120 haemophilus	363	30.5	7.2	89	1	YC36_METTH	P11697 staphylococ
291	31	7.3	75	1	MT1_CICAR	Q39458 cicier ariet	364	30.5	7.2	89	1	GG11_STAHA	Q19912 cyanidium c
292	31	7.3	75	1	YH11_CLOPE	Q8xjP5 clostridium	365	30.5	7.2	89	1	RG34_CVACA	Q9cdv5 lactococcus
293	31	7.3	76	1	SECG_MYCPN	Q9exd0 mycoplasma	366	30.5	7.2	89	1	R332_LACLA	P01078 solanum mel
294	31	7.3	76	1	TXP8_APTSC	P49270 apocostichus	367	30.5	7.2	89	1	IPR_SOLME	Q11116 caenorhabd
295	31	7.3	77	1	RL28_HAELN	P44364 haemophilus	368	30.5	7.2	89	1	IOVO_BALPA	P52239 balearica p
296	31	7.3	77	1	SECG_MYCE	P58061 mycoplasma	369	30.5	7.2	89	1	IOVO_HALIN	P05577 grus carunc
297	31	7.3	77	1	Y125_HUMAN	Q14138 homo sapien	370	30.5	7.2	89	1	IOVO_HALIN	P05577 grus carunc
298	31	7.3	78	1	M6_LILHE	P55409 rhizobium s	371	30.5	7.2	89	1	YC36_METTH	P05577 grus carunc
299	31	7.3	78	1	MULI_PROMI	Q40189 lilium henr	372	30.5	7.2	89	1	YC36_METTH	P05577 grus carunc
300	31	7.3	78	1	YD80_METJA	P09461 proteus mir	373	30.5	7.2	89	1	YC36_METTH	P05577 grus carunc
301	31	7.3	80	1	AKA7_MOUSE	Q58775 methanococc	374	30.5	7.2	89	1	YC36_METTH	P05577 grus carunc
302	31	7.3	80	1	MT2_FRAAN	O55074 mus musculu	375	30.5	7.2	89	1	YC36_METTH	P05577 grus carunc
303	31	7.3	81	1	NX52_AIPLA	P93134 alphysurus l	376	30.5	7.2	89	1	YC36_METTH	P05577 grus carunc
304	31	7.3	81	1	YF76_CLOPE	P19959 alphysurus l	377	30.5	7.2	89	1	YC36_METTH	P05577 grus carunc
305	31	7.3	81	1	YF76_CLOPE	Q8xjP4 clostridium	378	30.5	7.2	89	1	YC36_METTH	P05577 grus carunc
306	31	7.3	82	1	EXTS_RHIME	Q28696 archaeoglob	379	30.5	7.2	89	1	YC36_METTH	P05577 grus carunc
307	31	7.3	83	1	IMW5_ECOLI	Q921r9 rhizobium m	380	30.5	7.2	89	1	YC36_METTH	P05577 grus carunc
308	31	7.3	83	1	IMW5_ECOLI	P13476 escherichia	381	30.5	7.2	89	1	YC36_METTH	P05577 grus carunc
309	31	7.3	83	1	IMW5_ECOLI	Q57300 pseudomonas	382	30.5	7.2	89	1	YC36_METTH	P05577 grus carunc
310	31	7.3	83	1	MULI_PSEAE	P11221 pseudomonas	383	30.5	7.2	89	1	YC36_METTH	P05577 grus carunc
311	31	7.3	84	1	RNS_GIRCA	Q29535 giraffa cam	384	30.5	7.2	89	1	YC36_METTH	P05577 grus carunc
312	31	7.3	84	1	RS16_RALSO	Q8y0w0 ralstonia s	385	30.5	7.2	89	1	YC36_METTH	P05577 grus carunc
313	31	7.3	85	1	V12_BPT7	Q57755 methanococc	386	30.5	7.2	89	1	YC36_METTH	P05577 grus carunc
314	31	7.3	85	1	TAT_HV1H2	P03780 bacterioph	387	30.5	7.2	89	1	YC36_METTH	P05577 grus carunc
315	31	7.3	86	1	ANPE_NACAM	P04608 human immu	388	30.5	7.2	89	1	YC36_METTH	P05577 grus carunc
316	31	7.3	87	1	RS16_FUSNN	P19605 macrozoarce	389	30.5	7.2	89	1	YC36_METTH	P05577 grus carunc
317	31	7.3	87	1	EXTS_STRCO	Q8trv8 fusobacteri	390	30.5	7.2	89	1	YC36_METTH	P05577 grus carunc
318	31	7.3	88	1	YF76_CLOPE	Q9f5m4 streptomyc	391	30.5	7.2	89	1	YC36_METTH	P05577 grus carunc
319	31	7.3	88	1	YF76_CLOPE	P74891 salmonele	392	30.5	7.2	89	1	YC36_METTH	P05577 grus carunc
320	31	7.3	88	1	SSAS_SALTY	P48991 streptococ	393	30.5	7.2	89	1	YC36_METTH	P05577 grus carunc
321	31	7.3	88	1	VG30_BPMO2	Q64224 mycobacteri	394	30.5	7.2	89	1	YC36_METTH	P05577 grus carunc
322	31	7.3	90	1	MTFS_ECOLI	Q46971 escherichia	395	30.5	7.2	89	1	YC36_METTH	P05577 grus carunc
323	31	7.3	90	1	X36A_SOYBN	Q02918 glycine max	396	30.5	7.2	89	1	YC36_METTH	P05577 grus carunc
324	31	7.3	90	1	RP0Z_VIBCH	Q9knm3 vibrio chol	397	30.5	7.2	89	1	YC36_METTH	P05577 grus carunc
325	30.5	7.2	45	1	VPU_HV1N5	P08804 human immu	398	30.5	7.2	89	1	YC36_METTH	P05577 grus carunc



399	7.1	71	1	BDM_ECOLI	P76127	escherichia	472	29.5	7.0	80	1	PSAC_FREDI	P23810	fremyella d
400	30	7.1	1	NTB2_XENLA	P19011	xenopus lae	473	29.5	7.0	80	1	PSAC_PORPU	P51374	porphyra pu
401	30	7.1	72	NTK4_OPHPA	P80156	ophiophagus	474	29.5	7.0	80	1	PSAC_SINY3	P32422	synochocyst
402	30	7.1	72	NXL5_OPHPA	P80965	ophiophagus	475	29.5	7.0	80	1	S61G_YEAST	P35179	saccharomyc
403	30	7.1	73	SAS5_BACME	P04835	bacillus me	476	29.5	7.0	81	1	TXFB_DENAN	P01404	dendroaspis
404	30	7.1	74	COLE_ZOPAT	P80032	zophobas at	477	29.5	7.0	81	1	VPU_HVLY2	P35966	human immun
405	30	7.1	74	TXPI_APTSC	P49367	apostichus	478	29.5	7.0	82	1	NXSB_NAJAT	P80958	naja atra (
406	30	7.1	75	YH56_CLOAB	Q97196	ciostriidum	479	29.5	7.0	82	1	YFDX_PLEBO	P46043	plectonema
407	30	7.1	76	SPY1_CHICK	Q97411	gallus gall	480	29.5	7.0	84	1	IAC2_HUMAN	P20155	homo sapien
408	30	7.1	76	TEC_BFPI1	Q06461	bacteriophia	481	29.5	7.0	84	1	RL24_SULAC	O05633	sulfolobus
409	30	7.1	77	CSME_CHLTE	O88986	chlorobium	482	29.5	7.0	86	1	PAK1_SINY3	P72712	synochocyst
410	30	7.1	77	LEA2_PIG	Q95104	sus scrofa	483	29.5	7.0	86	1	SCX1_CENSC	F01491	centruroides
411	30	7.1	77	YF67_XYLFA	Q99933	xyella fas	484	29.5	7.0	87	1	SSS1_SCICA	P13275	scyllorhinu
412	30	7.1	77	YHR2_VACCV	P17358	xylella vi	485	29.5	7.0	88	1	PHHP_LISMO	O31148	listeria mo
413	30	7.1	78	DLTC_STAAM	O53663	staphylococ	486	29.5	7.0	88	1	YPER4_NPVLD	P36868	lymantria d
414	30	7.1	78	DLTC_STAAX	O9x2n6	staphylococ	487	29.5	7.0	89	1	PERC_ECO27	P43475	escherichia
415	30	7.1	78	R28B_STRCO	O9x8k8	streptomyce	488	29.5	7.0	90	1	LITP_PHYSA	P08947	phyliomedus
416	30	7.1	79	EX7S_LACLA	O9x8k8	lactococcus	489	29.5	7.0	90	1	R37A_OSTOS	O61598	ostertagia
417	30	7.1	80	MT2_BRARA	Q42494	brassica ra	490	29.5	7.0	90	1	RS16_LACLA	Q9cib2	lactococcus
418	30	7.1	80	MT2_BRARP	Q39269	brassica ra	491	29.5	6.9	43	1	CC1_CARCN	P32954	carica cand
419	30	7.1	80	MT2_RICCO	P30564	ricinus com	492	29.5	6.9	44	1	DERM_PIG	P45846	sus scrofa
420	30	7.1	80	XJBE_ECOLI	P32686	escherichia	493	29.5	6.9	46	1	RL34_PSEAO	O86448	pseudanabae
421	30	7.1	80	VV6A_VIBCH	P58093	vibrio chol	494	29.5	6.9	47	1	RL40_METJA	P54058	methanococ
422	30	7.1	81	ATPH_PEA	P08212	pisum sativ	495	29.5	6.9	48	1	RK32_VICFA	P15820	vicia faba
423	30	7.1	81	ATPH_SPIOL	P00843	spinacia ol	496	29.5	6.9	49	1	YLDH_MYCHY	P46188	mycoplasma
424	30	7.1	81	PER_BAGST	P00212	baginellus st	497	29.5	6.9	50	1	INS_PROGU	P01331	proechimys
425	30	7.1	82	RS16_YERPE	O82b07	yersinia pe	498	29.5	6.9	50	1	RS14_METH	O36125	methanobact
426	30	7.1	83	CNCG_SFEIR	O55175	spemophilu	499	29.5	6.9	51	1	LOVO_RHYFU	P52251	rhynchobut
427	30	7.1	83	ELAC_MACEU	O62845	macropus eu	500	29.5	6.9	51	1	LHB2_ECTHA	P80105	ectothiorho
428	30	7.1	83	VP19_BRAPS	Q9t1q9	bacteriophia	501	29.5	6.9	54	1	TOVO_GYECO	P05578	gyps coprot
429	30	7.1	84	RS14_ARATH	P42036	arabidopsis	502	29.5	6.9	54	1	TOVO_HALLA	P52268	hallaetetus
430	30	7.1	85	GLRL_ECOLI	P00277	escherichia	503	29.5	6.9	55	1	ELAG_ADE05	P06438	humandeno
431	30	7.1	85	IATP_YEAST	P01097	saccharomyc	504	29.5	6.9	55	1	RA32_SPIOL	P06804	spinnacia ol
432	30	7.1	85	YME1_THIFE	P22904	thiobacillu	505	29.5	6.9	56	1	BEI_HUMAN	P02090	homo sapien
433	30	7.1	86	ACP_CYACA	O19921	cyanidium c	506	29.5	6.9	57	1	FER_DESGI	Q09401	caenorhabdi
434	30	7.1	86	RS18_CAMJE	O69301	campylobact	507	29.5	6.9	58	1	YRD5_CABEL	O85559	archaeoglob
435	30	7.1	86	SCX1_CENNO	P15223	centruroides	508	29.5	6.9	59	1	YH14_ARCFU	P07215	saccharomyc
436	30	7.1	86	YABA_ESCFE	P45665	centruroides	509	29.5	6.9	60	1	MYC_YEAST	O83797	treponema p
437	30	7.1	86	YALB_KLEAE	P28243	escherichia	510	29.5	6.9	61	1	Y825_TREPA	O93173	archaeoglob
438	30	7.1	87	SELW_MOUSE	P51414	arabidopsis	511	29.5	6.9	61	1	YA92_ARCFU	O88200	vicia faba
439	30	7.1	88	RS16_METAC	P49904	mus musculus	512	29.5	6.9	62	1	ICIS_VICFA	P15113	candida gla
440	30	7.1	88	YV7A_AERPE	Q88q97	mycoplasma	513	29.5	6.9	62	1	MT1_CANGA	P10458	laticauda c
441	30	7.1	89	RS16_BACSU	P21474	bacillus su	514	29.5	6.9	62	1	MXSC_LATCR	P10460	laticauda l
442	30	7.1	89	Y27A_AERPE	P58322	aeropyrum p	515	29.5	6.9	62	1	YPTA_BACSL	P34521	caenorhabdi
443	30	7.1	89	YVAX_KLEAE	Q97783	mus musculus	517	29.5	6.9	63	1	YPTA_BACSL	P50841	caenorhabdi
444	30	7.1	90	AFIO_MOUSE	P97783	mus musculus	518	29.5	6.9	63	1	ANP2_MACAM	P19608	macrozoarce
445	30	7.1	90	ASIA_BPT4	P32267	bacteriophia	519	29.5	6.9	64	1	RL35_CHLTR	O84841	chlamydia t
446	30	7.1	90	VPM_BPPRD	P27389	bacteriophia	520	29.5	6.9	64	1	Y64_LAMBD	P24331	crotalus du
447	30	7.1	90	YAVE_SCHPO	Q10189	schizosacch	521	29.5	6.9	65	1	MX1_CRODU	P24331	crotalus du
448	30	7.1	90	GIP_BOVIN	P09680	bos taurus	522	29.5	6.9	65	1	MX3_CRODU	P95669	thermococcu
449	29.5	7.0	47	PEG_SYNP6	O8079	synochococ	523	29.5	6.9	66	1	HANA_THEZI	Q17313	ceratitidis c
450	29.5	7.0	47	UCR3_DROME	O9XV35	drosophila	524	29.5	6.9	66	1	CERC_CERCA	Q17313	haemophilus
451	29.5	7.0	55	CD52_MACFA	P32763	macaca fasc	525	29.5	6.9	67	1	YEO2_HAEIN	Q04646	mus musculus
452	29.5	7.0	60	CD52_MACFA	P32763	macaca fasc	526	29.5	6.9	68	1	ATNG_MOUSE	P16221	rattus norv
453	29.5	7.0	61	CD52_HUMAN	P32763	macaca fasc	527	29.5	6.9	70	1	COXQ_RAT	P78977	yersinia en
454	29.5	7.0	61	YW98_PIRAB	Q9uz10	pyrococcus	528	29.5	6.9	71	1	HSTB_FEREN	P78699	kluyveromyc
455	29.5	7.0	63	SCXV_CENSC	P46066	centruroides	529	29.5	6.9	71	1	SOMI_KLULA	Q09005	xenopus lae
456	29.5	7.0	63	Y008_BPL2	P42543	bacteriophia	530	29.5	6.9	71	1	STIB_XENLA	P19009	xenopus lae
457	29.5	7.0	66	RPB1_CAEFR	P35074	caenorhabdi	531	29.5	6.9	71	1	VITAL_XENLA	Q9pdd4	xylella fas
458	29.5	7.0	66	YH18_ARCFU	O38556	archaeoglob	532	29.5	6.9	72	1	IF1_XYLFA	Q9pdd4	xylella fas
459	29.5	7.0	67	HMTB_PYRKO	Q9y812	pyrococcus	533	29.5	6.9	72	1	RPOZ_CLOAB	Q38625	bacteriophia
460	29.5	7.0	67	HMTB_METH	P50484	methanobact	534	29.5	6.9	72	1	VG18_BPMU	Q38625	bacteriophia
461	29.5	7.0	67	YVAV_VACCV	P20528	vaccinia vi	535	29.5	6.9	72	1	NXLI_NAJOX	P01382	naja oxiata
462	29.5	7.0	68	YB10_HALNI	P17104	halobacteri	536	29.5	6.9	73	1	RL24_HELPJ	Q9zjs3	helicobacte
463	29.5	7.0	68	GRGU_HUMAN	O14610	homo sapien	537	29.5	6.9	73	1	DNBI_POVBA	P14998	polymaviru
464	29.5	7.0	73	RS19_YERPS	P11265	yersinia ps	538	29.5	6.9	74	1	VIAL_MYCTU	P19771	mycobacteri
465	29.5	7.0	75	EX7S_LISIN	Q92bz2	listeria in	539	29.5	6.9	74	1	Z123_HUMAN	P35273	homo sapien
466	29.5	7.0	75	EX7S_LISIN	O8y7c3	listeria mo	540	29.5	6.9	75	1	GBG4_MOUSE	P50150	homo sapien
467	29.5	7.0	75	YML1_THIFE	P20087	thiobacillu	541	29.5	6.9	75	1	Y13H_BPT4	P39501	bacteriophia
468	29.5	7.0	76	RI16_GUTH	O78423	quillardia	542	29.5	6.9	75	1	YB7F_SCHPO	O14330	schizosacch
469	29.5	7.0	77	YVFB_VACCC	P14214	tachypleus	543	29.5	6.9	75	1	Y13H_BPT4	O14330	schizosacch
470	29.5	7.0	77	YVFB_VACCC	P14214	tachypleus	543	29.5	6.9	75	1	Y13H_BPT4	O14330	schizosacch
471	29.5	7.0	80	C550_THAWE	O98448	thalassios	544	29.5	6.9	75	1	Y13H_BPT4	O14330	schizosacch

545	29	6.9	75	1	VE69_METJA	Q58864	methanococc	618	28.5	6.7	68	1	MTPE_METJA	Q58262	methanococc
546	29	6.9	75	1	VEED_ECOLI	P33014	escherichia	619	28.5	6.7	68	1	YRP3_AZOCA	P33988	azorhizobiu
547	29	6.9	76	1	ETL_HUMAN	Q24869	macaca fasc	620	28.5	6.7	69	1	CSPA_RHIME	Q92386	rhizobium m
548	29	6.9	76	1	RUXG_HUMAN	Q15357	homo sapien	621	28.5	6.7	70	1	CCRN_TRASC	P80345	trachemys s
549	29	6.9	76	1	YNGJ_ECOLI	P76105	escherichia	622	28.5	6.7	70	1	CSPA_VIBCH	Q9kn00	vibrio chol
550	29	6.9	77	1	YQAF_BACSU	P45903	bacillus su	623	28.5	6.7	70	1	Y650_HAEIN	P44028	haemophilus
551	29	6.9	77	1	LEA2_WACMU	Q95903	macaca mulia	624	28.5	6.7	71	1	NIFU_FRAAL	P46045	frankia aln
552	29	6.9	77	1	YDCE_ECOLI	P31262	escherichia	625	28.5	6.7	71	1	YC47_PORPU	P51274	porphyra pu
553	29	6.9	77	1	YDCE_ECOLI	Q28636	archaeoglob	626	28.5	6.7	72	1	YDGR_ECOLI	Q46932	escherichia
554	29	6.9	77	1	MTJ2_NTCGU	Q40396	nicotiana g	627	28.5	6.7	72	1	ATX1_YEAST	P38636	saccharomyc
555	29	6.9	78	1	MTJ8_TREPA	P96131	treponema p	628	28.5	6.7	73	1	H14_RABIT	P22522	oryctolagus
556	29	6.9	78	1	RT19_ACACA	P46762	acanthamoeb	629	28.5	6.7	73	1	VF07_VACCV	P24359	vaccinia vi
557	29	6.9	79	1	AFI3_BRANA	Q39313	brassica na	630	28.5	6.7	77	1	MTIA_HSV6U	Q41669	vicia faba
558	29	6.9	79	1	YF71_HAEIN	P44260	haemophilus	631	28.5	6.7	77	1	RI16_CYPAP	P48139	cyanophora
559	29	6.9	80	1	ATPD_FIG	Q95312	sus scrofa	632	28.5	6.7	77	1	UL11_HSV6U	P24448	human herpe
560	29	6.9	80	1	VPI_BIV27	P24035	bovine immu	633	28.5	6.7	77	1	Y948_HAEIN	Q57120	haemophilus
561	29	6.9	80	1	YK09_MYCTU	Q10848	mycobacteri	634	28.5	6.7	80	1	EX7S_BACHD	Q9K568	bacillus ha
562	29	6.9	82	1	ATPH_CHURE	Q37304	chlamydomon	635	28.5	6.7	80	1	EX7S_RICCN	Q921e5	rickettsia
563	29	6.9	82	1	MTJ2_ORYSA	P94029	oryza sativ	636	28.5	6.7	80	1	PSAC_ANASP	P23392	anabaena sp
564	29	6.9	83	1	RL23_HALNI	Q06842	halobacteri	637	28.5	6.7	80	1	PSAC_CHLVU	P56301	chlorocella v
565	29	6.9	83	1	RL23_HALNI	P51305	porphyra pu	638	28.5	6.7	80	1	PSAC_EUGER	P31556	euglena gra
566	29	6.9	83	1	RR17_PORPU	P42677	homo sapien	639	28.5	6.7	80	1	PSAC_MAYZE	P11061	zea mays (m
567	29	6.9	83	1	RS27_HUMAN	P24051	rattus norv	640	28.5	6.7	80	1	PSAC_MARFO	P06251	marichantia
568	29	6.9	83	1	RS27_XENLA	P47904	xenopus lae	641	28.5	6.7	80	1	PSAC_ORISA	P10794	oryza sativ
569	29	6.9	83	1	Y46S_SYNY3	P73882	synchocyst	642	28.5	6.7	80	1	PSAC_PSINU	P58870	psilotum nu
570	29	6.9	84	1	IM13_ORYSA	Q9XGVS	oryza sativ	643	28.5	6.7	80	1	PSAC_SYNEP	P18083	synchococc
571	29	6.9	84	1	SC13_MESMA	Q9NJC8	mesobuthus	644	28.5	6.7	80	1	PSAC_SYNEP	P31087	synchococc
572	29	6.9	85	1	HL15_MAIZE	P56663	zea mays (m	645	28.5	6.7	80	1	PSAC_SYNEP	Q69195	synchococc
573	29	6.9	85	1	HLG3_MAIZE	P56669	zea mays (m	646	28.5	6.7	80	1	VF07_VACCP	P29886	vaccinia vi
574	29	6.9	85	1	PS17_PREDI	P11397	iremyella d	647	28.5	6.7	81	1	VPU_HVIEL	P05925	human immun
575	29	6.9	85	1	PS17_PREDI	Q95WDL	mycoplasma	648	28.5	6.7	81	1	VPU_HV122	P18806	human immun
576	29	6.9	85	1	SCX8_CENSC	Q2570	streptomyce	649	28.5	6.7	81	1	XENO_XENLA	P07198	xenopus lae
577	29	6.9	86	1	AGX8_STRCN	Q02570	streptomyce	650	28.5	6.7	81	1	Y697_RICPR	Q9ZCM7	rickettsia
578	29	6.9	86	1	MYSE_DICDI	P54695	dictyosteli	651	28.5	6.7	82	1	TOLB_ACTPL	Q44156	actinobacil
579	29	6.9	86	1	YFHL_HAEIN	P47476	haemophilus	652	28.5	6.7	84	1	VG35_HAEIN	P44228	haemophilus
580	29	6.9	87	1	ANF5_MACAM	P19607	macrozoarce	653	28.5	6.7	85	1	YSH7_CABEL	P50438	caenorhabdi
581	29	6.9	87	1	ANPD_MACAM	P19604	macrozoarce	654	28.5	6.7	85	1	RL27_MYCTU	P71908	mycobacteri
582	29	6.9	87	1	RS20_CAMUE	Q9PM64	campylobact	655	28.5	6.7	86	1	ATPE_CHLLI	P35111	chloridium
583	29	6.9	87	1	SCX1_CLOPE	P01492	centruroid	656	28.5	6.7	88	1	MINE_GUITH	Q78435	guillardia
584	29	6.9	87	1	SCX1_CENSC	P01492	centruroid	657	28.5	6.7	88	1	PTHF_LACSK	Q07125	lactobacill
585	29	6.9	87	1	Y230_CENNO	Q57683	methanococc	658	28.5	6.7	88	1	Y002_NPVOP	O10355	orgyia pseu
586	29	6.9	87	1	Y230_CENNO	P58235	synchocyst	659	28.5	6.7	88	1	YB1T_ECOLI	P11039	escherichia
587	29	6.9	88	1	ELB9_ADE07	P04490	human adeno	660	28.5	6.7	89	1	YB1T_ECOLI	P11540	chlamydia am
588	29	6.9	88	1	FLGM_BACSU	P39809	bacillus su	661	28.5	6.7	89	1	RL28_CHLMU	Q9PKV0	chlamydia t
589	29	6.9	88	1	FLGM_BACSU	P56668	zea mays (m	662	28.5	6.7	89	1	RL28_CHLTR	O84088	chlamydia t
590	29	6.9	88	1	HLKB_MAIZE	Q97TC9	streptococc	663	28.5	6.7	89	1	RS15_MYCTU	O33327	mycobacteri
591	29	6.9	88	1	Y007_STRPN	P03853	escherichia	664	28.5	6.7	89	1	YB16_MCGE	P47684	mycoplasma
592	29	6.9	88	1	YB4_ECOLI	Q91G01	neisseria m	665	28.5	6.7	89	1	YB16_MCGE	O28885	archaeoglob
593	29	6.9	89	1	PHPP_NEIMA	P55563	rhizobium s	666	28.5	6.7	90	1	IM10_MOUSE	Q9YVW0	mus musculu
594	29	6.9	89	1	YAMD_RHISN	P55563	rhizobium s	667	28.5	6.7	90	1	Y060_NPVOP	O10317	orgyia pseu
595	29	6.9	90	1	DBH_HAEIN	P43722	haemophilus	668	28.5	6.7	90	1	Y060_NPVOP	P06849	wheat dwarf
596	29	6.9	90	1	TATP_SCHPO	O74523	schizosacch	669	28.5	6.7	90	1	Y060_NPVOP	O10317	orgyia pseu
597	29	6.9	90	1	R35A_METKA	Q8TV66	methanopyru	670	28.5	6.7	90	1	Y060_NPVOP	O10317	orgyia pseu
598	29	6.9	90	1	RS16_STRPY	P58124	streptococc	671	28.5	6.7	90	1	Y060_NPVOP	O10317	orgyia pseu
599	29	6.9	90	1	VP49_BRAPS	Q8TV66	methanopyru	672	28.5	6.7	90	1	Y060_NPVOP	O10317	orgyia pseu
600	29	6.9	90	1	R332_BACSU	Q8TV66	methanopyru	673	28.5	6.7	90	1	Y060_NPVOP	O10317	orgyia pseu
601	28.5	6.7	50	1	BA11_SCHCO	P56504	bacillus su	674	28.5	6.7	90	1	Y060_NPVOP	O10317	orgyia pseu
602	28.5	6.7	50	1	Y05F_PIG	Q29288	sus scrofa	675	28.5	6.7	90	1	Y060_NPVOP	O10317	orgyia pseu
603	28.5	6.7	52	1	CD3E_BOVIN	Q28073	bos taurus	676	28.5	6.7	90	1	Y060_NPVOP	O10317	orgyia pseu
604	28.5	6.7	53	1	LEC2_LATOC	P12306	lathyrus oc	677	28.5	6.7	90	1	Y060_NPVOP	O10317	orgyia pseu
605	28.5	6.7	53	1	LEC2_LATOC	P12307	lathyrus oc	678	28.5	6.7	90	1	Y060_NPVOP	O10317	orgyia pseu
606	28.5	6.7	53	1	LECA_LATAR	P07442	lathyrus ar	679	28.5	6.7	90	1	Y060_NPVOP	O10317	orgyia pseu
607	28.5	6.7	56	1	RS14_PYRHO	O74093	pyrococcus	680	28.5	6.7	90	1	Y060_NPVOP	O10317	orgyia pseu
608	28.5	6.7	57	1	RL32_MYCLE	Q9CD69	mycobacteri	681	28.5	6.7	90	1	Y060_NPVOP	O10317	orgyia pseu
609	28.5	6.7	60	1	Y569_ARCFU	O19199	archaeoglob	682	28.5	6.7	90	1	Y060_NPVOP	O10317	orgyia pseu
610	28.5	6.7	63	1	HIG2_HUMAN	O9Y512	homo sapien	683	28.5	6.7	90	1	Y060_NPVOP	O10317	orgyia pseu
611	28.5	6.7	65	1	TXMW_NAJKA	P82935	naja najka	684	28.5	6.7	90	1	Y060_NPVOP	O10317	orgyia pseu
612	28.5	6.7	66	1	COPP_HELPF	Q32620	helicobacte	685	28.5	6.7	90	1	Y060_NPVOP	O10317	orgyia pseu
613	28.5	6.7	66	1	RCPP_LAMBD	P03040	bacterioph	686	28.5	6.7	90	1	Y060_NPVOP	O10317	orgyia pseu
614	28.5	6.7	66	1	Y2AB_PSEAE	O51384	pseudomonas	687	28.5	6.7	90	1	Y060_NPVOP	O10317	orgyia pseu
615	28.5	6.7	66	1	YSCB_YREAN	O1246	persinia en	688	28.5	6.7	90	1	Y060_NPVOP	O10317	orgyia pseu
616	28.5	6.7	67	1	Y158_RICPR	Q9Z404	rickettsia	689	28.5	6.7	90	1	Y060_NPVOP	O10317	orgyia pseu
617	28.5	6.7	68	1	ITH5_CUCMA	P19873	cucurbita m	690	28.5	6.7	90	1	Y060_NPVOP	O10317	orgyia pseu

691	28	6.6	57	1	IVBK_DENPO	P00981 dendroaspis	764	28	6.6	80	1	PSAC_MESVI	Q9mum9 mesostigma
692	28	6.6	57	1	NIND_LAMED	P03767 bacterioph	765	28	6.6	80	1	PSAC_NEPOL	Q9tkv9 nephroselmi
693	28	6.6	58	1	RHO_THIFE	P52158 thiodacillu	766	28	6.6	80	1	RS16_MYCGA	Q9rdv4 mycoplasma
694	28	6.6	58	1	TX50_DENJA	P01419 dendroaspis	767	28	6.6	81	1	ATPH_ARATH	P56760 arabidopsis
695	28	6.6	58	1	V3A_IBVB	P30237 avian infec	768	28	6.6	82	1	ATPH_GUITH	O78479 guillardia
696	28	6.6	59	1	Y0AH_SALRY	P56505 salmonella	769	28	6.6	82	1	CTJA_BACSU	O45536 bacillus su
697	28	6.6	60	1	MERC_SHIFL	P04337 shigella fl	770	28	6.6	82	1	CXOB_CONST	O9x2k5 conus stria
698	28	6.6	60	1	NXS1_ENHSC	P25492 enhydrina s	771	28	6.6	82	1	IACP_SALTI	O8z495 salmonella
699	28	6.6	60	1	NXS1_HYDCY	P25494 enhydrina s	772	28	6.6	82	1	IMCL_SALTI	P18643 salmoneilla
700	28	6.6	60	1	NXS2_ENHSC	P25493 enhydrina s	773	28	6.6	82	1	SASG_BACST	P076199 phenetria st
701	28	6.6	61	1	NXS3_NAJHA	P01420 najja haje a	774	28	6.6	82	1	UXSA_PIG	O76199 phenetria st
702	28	6.6	61	1	RL4A_MYCTU	P95065 mycobacteri	775	28	6.6	82	1	UXSA_PHONI	O29150 sus scrofa
703	28	6.6	62	1	IT13_MOMCH	P09407 momordica c	776	28	6.6	82	1	YAS7_VIBCH	O9k448 vibrio chol
704	28	6.6	62	1	P29789 PORPHYRA PU	P29789 porphyra pu	777	28	6.6	82	1	YCXL_VICFA	P08889 victia faba
705	28	6.6	62	1	SCX2_BUTEU	P15221 buthus eupe	778	28	6.6	83	1	NSGL_RAT	P02883 rattus norv
706	28	6.6	62	1	YPPE_BACSU	P50834 bacillus su	779	28	6.6	83	1	RS28_SULTO	Q97328 sulfolobus
707	28	6.6	63	1	COXO_MOUSE	P17665 mus musculu	780	28	6.6	83	1	RUXX_THEAC	P57670 thermoplas
708	28	6.6	63	1	LANN_SPMU	P22287 cladosporiu	781	28	6.6	83	1	RUXX_THEAC	Q10703 mycobacteri
709	28	6.6	63	1	YL33_DROME	O68586 streptococc	782	28	6.6	83	1	TATA_MYCTU	Q05242 mycobacteri
710	28	6.6	63	1	YL33_DROME	Q9vr93 rhodobacter	783	28	6.6	83	1	VG03_BPML5	P27391 bacterioph
711	28	6.6	63	1	YL33_DROME	P30791 rhodobacter	784	28	6.6	83	1	VPP_BPPRD	O9ze49 rickettsia
712	28	6.6	64	1	YL33_DROME	O9ze49 rickettsia	785	28	6.6	83	1	Y098_RICPR	Q9x713 rhizobium m
713	28	6.6	64	1	Y098_RICPR	O9ze49 rickettsia	786	28	6.6	84	1	Y098_RICPR	P12049 bacillus su
714	28	6.6	64	1	Y098_RICPR	O9ze49 rickettsia	787	28	6.6	84	1	Y098_RICPR	Q9pdq9 xylella fas
715	28	6.6	64	1	Y098_RICPR	O9ze49 rickettsia	788	28	6.6	85	1	Y098_RICPR	Q9pdq9 xylella fas
716	28	6.6	64	1	Y098_RICPR	O9ze49 rickettsia	789	28	6.6	85	1	Y098_RICPR	Q9pdq9 xylella fas
717	28	6.6	64	1	Y098_RICPR	O9ze49 rickettsia	790	28	6.6	85	1	Y098_RICPR	Q9pdq9 xylella fas
718	28	6.6	64	1	Y098_RICPR	O9ze49 rickettsia	791	28	6.6	85	1	Y098_RICPR	Q9pdq9 xylella fas
719	28	6.6	64	1	Y098_RICPR	O9ze49 rickettsia	792	28	6.6	85	1	Y098_RICPR	Q9pdq9 xylella fas
720	28	6.6	64	1	Y098_RICPR	O9ze49 rickettsia	793	28	6.6	85	1	Y098_RICPR	Q9pdq9 xylella fas
721	28	6.6	64	1	Y098_RICPR	O9ze49 rickettsia	794	28	6.6	85	1	Y098_RICPR	Q9pdq9 xylella fas
722	28	6.6	64	1	Y098_RICPR	O9ze49 rickettsia	795	28	6.6	85	1	Y098_RICPR	Q9pdq9 xylella fas
723	28	6.6	64	1	Y098_RICPR	O9ze49 rickettsia	796	28	6.6	85	1	Y098_RICPR	Q9pdq9 xylella fas
724	28	6.6	64	1	Y098_RICPR	O9ze49 rickettsia	797	28	6.6	85	1	Y098_RICPR	Q9pdq9 xylella fas
725	28	6.6	64	1	Y098_RICPR	O9ze49 rickettsia	798	28	6.6	85	1	Y098_RICPR	Q9pdq9 xylella fas
726	28	6.6	64	1	Y098_RICPR	O9ze49 rickettsia	799	28	6.6	85	1	Y098_RICPR	Q9pdq9 xylella fas
727	28	6.6	64	1	Y098_RICPR	O9ze49 rickettsia	800	28	6.6	85	1	Y098_RICPR	Q9pdq9 xylella fas
728	28	6.6	64	1	Y098_RICPR	O9ze49 rickettsia	801	28	6.6	85	1	Y098_RICPR	Q9pdq9 xylella fas
729	28	6.6	64	1	Y098_RICPR	O9ze49 rickettsia	802	28	6.6	85	1	Y098_RICPR	Q9pdq9 xylella fas
730	28	6.6	64	1	Y098_RICPR	O9ze49 rickettsia	803	28	6.6	85	1	Y098_RICPR	Q9pdq9 xylella fas
731	28	6.6	64	1	Y098_RICPR	O9ze49 rickettsia	804	28	6.6	85	1	Y098_RICPR	Q9pdq9 xylella fas
732	28	6.6	64	1	Y098_RICPR	O9ze49 rickettsia	805	28	6.6	85	1	Y098_RICPR	Q9pdq9 xylella fas
733	28	6.6	64	1	Y098_RICPR	O9ze49 rickettsia	806	28	6.6	85	1	Y098_RICPR	Q9pdq9 xylella fas
734	28	6.6	64	1	Y098_RICPR	O9ze49 rickettsia	807	28	6.6	85	1	Y098_RICPR	Q9pdq9 xylella fas
735	28	6.6	64	1	Y098_RICPR	O9ze49 rickettsia	808	28	6.6	85	1	Y098_RICPR	Q9pdq9 xylella fas
736	28	6.6	64	1	Y098_RICPR	O9ze49 rickettsia	809	28	6.6	85	1	Y098_RICPR	Q9pdq9 xylella fas
737	28	6.6	64	1	Y098_RICPR	O9ze49 rickettsia	810	28	6.6	85	1	Y098_RICPR	Q9pdq9 xylella fas
738	28	6.6	64	1	Y098_RICPR	O9ze49 rickettsia	811	28	6.6	85	1	Y098_RICPR	Q9pdq9 xylella fas
739	28	6.6	64	1	Y098_RICPR	O9ze49 rickettsia	812	28	6.6	85	1	Y098_RICPR	Q9pdq9 xylella fas
740	28	6.6	64	1	Y098_RICPR	O9ze49 rickettsia	813	28	6.6	85	1	Y098_RICPR	Q9pdq9 xylella fas
741	28	6.6	64	1	Y098_RICPR	O9ze49 rickettsia	814	28	6.6	85	1	Y098_RICPR	Q9pdq9 xylella fas
742	28	6.6	64	1	Y098_RICPR	O9ze49 rickettsia	815	28	6.6	85	1	Y098_RICPR	Q9pdq9 xylella fas
743	28	6.6	64	1	Y098_RICPR	O9ze49 rickettsia	816	28	6.6	85	1	Y098_RICPR	Q9pdq9 xylella fas
744	28	6.6	64	1	Y098_RICPR	O9ze49 rickettsia	817	28	6.6	85	1	Y098_RICPR	Q9pdq9 xylella fas
745	28	6.6	64	1	Y098_RICPR	O9ze49 rickettsia	818	28	6.6	85	1	Y098_RICPR	Q9pdq9 xylella fas
746	28	6.6	64	1	Y098_RICPR	O9ze49 rickettsia	819	28	6.6	85	1	Y098_RICPR	Q9pdq9 xylella fas
747	28	6.6	64	1	Y098_RICPR	O9ze49 rickettsia	820	28	6.6	85	1	Y098_RICPR	Q9pdq9 xylella fas
748	28	6.6	64	1	Y098_RICPR	O9ze49 rickettsia	821	28	6.6	85	1	Y098_RICPR	Q9pdq9 xylella fas
749	28	6.6	64	1	Y098_RICPR	O9ze49 rickettsia	822	28	6.6	85	1	Y098_RICPR	Q9pdq9 xylella fas
750	28	6.6	64	1	Y098_RICPR	O9ze49 rickettsia	823	28	6.6	85	1	Y098_RICPR	Q9pdq9 xylella fas
751	28	6.6	64	1	Y098_RICPR	O9ze49 rickettsia	824	28	6.6	85	1	Y098_RICPR	Q9pdq9 xylella fas
752	28	6.6	64	1	Y098_RICPR	O9ze49 rickettsia	825	28	6.6	85	1	Y098_RICPR	Q9pdq9 xylella fas
753	28	6.6	64	1	Y098_RICPR	O9ze49 rickettsia	826	28	6.6	85	1	Y098_RICPR	Q9pdq9 xylella fas
754	28	6.6	64	1	Y098_RICPR	O9ze49 rickettsia	827	28	6.6	85	1	Y098_RICPR	Q9pdq9 xylella fas
755	28	6.6	64	1	Y098_RICPR	O9ze49 rickettsia	828	28	6.6	85	1	Y098_RICPR	Q9pdq9 xylella fas
756	28	6.6	64	1	Y098_RICPR	O9ze49 rickettsia	829	28	6.6	85	1	Y098_RICPR	Q9pdq9 xylella fas
757	28	6.6	64	1	Y098_RICPR	O9ze49 rickettsia	830	28	6.6	85	1	Y098_RICPR	Q9pdq9 xylella fas
758	28	6.6	64	1	Y098_RICPR	O9ze49 rickettsia	831	28	6.6	85	1	Y098_RICPR	Q9pdq9 xylella fas
759	28	6.6	64	1	Y098_RICPR	O9ze49 rickettsia	832	28	6.6	85	1	Y098_RICPR	Q9pdq9 xylella fas
760	28	6.6	64	1	Y098_RICPR	O9ze49 rickettsia	833	28	6.6	85	1	Y098_RICPR	Q9pdq9 xylella fas
761	28	6.6	64	1	Y098_RICPR	O9ze49 rickettsia	834	28	6.6	85	1	Y098_RICPR	Q9pdq9 xylella fas
762	28	6.6	64	1	Y098_RICPR	O9ze49 rickettsia	835	28	6.6	85	1	Y098_RICPR	Q9pdq9 xylella fas
763	28	6.6	64	1	Y098_RICPR	O9ze49 rickettsia	836	28	6.6	85	1	Y098_RICPR	Q9pdq9 xylella fas



## RESULT 2

Y16K_BPT4	STANDARD;	PRT;	71 AA.
ID_Y16K_BPT4			

IN  
RP. SEQUENCE FROM N.A.

```

CC SIGNAL TRANSDUCER CAN PLAY VERY DIFFERENT ROLES IN DIFFERENT CELL
CC TYPES AS A DIRECT CONSEQUENCE OF ITS GLYCOSYLATION.
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC -!- TISSUE SPECIFICITY: EXPRESSED IN THE CENTRAL NERVOUS SYSTEM, IN
CC POSTMITOTIC CELLS OF SPINAL CORD, HINDBRAIN, MIDBRAIN AND
CC FOREBRAIN. EXPRESSED IN EPITHELIUM DURING THE DEVELOPMENT OF NON-
CC NEURAL TISSUES. EXPRESSED IN TOOTH DEVELOPMENT, SPECIFICALLY IN
CC MESENCHYMAL CELLS DIFFERENTIATING INTO ODONTOBLAST IN DENTAL
CC PAPILLA, AS WELL AS IN THE DEVELOPING EYE AND HAIR FOLLICLE.
CC -!- DEVELOPMENTAL STAGE: DETECTED IN PRIMITIVE ECTODERM, MESODERM AND
CC -!- VENTRAL ENDODERM; DOWN-REGULATED WHEN ORGANOGENESIS IS COMPLETED.
CC -!- PTM: EXTENSIVELY O-GLYCOSYLATED (BY SIMILARITY). THE CARBOHYDRATE
CC STRUCTURE MAY BE REGULATED IN A TISSUE SPECIFIC AND DEVELOPMENTAL
CC STAGE SPECIFIC MANNER.
CC -!- SIMILARITY: TO OTHER MAMMALIAN SPECIES CD24.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC
CC EMBL; Z11663; CAA77731.1; -
CC GENE;
CC EMBL; U49062; AAG91470.1; -
CC GLYCOPROTEIN; GPI-anchor; Membrane; Signal; Differentiation.
CC FT SIGNAL 1 26 POTENTIAL.
CC FT CHAIN 27 56 SIGNAL TRANSDUCER CD24.
CC FT PROPEP 57 76 REMOVED IN MATURE FORM (BY SIMILARITY).
CC FT CARBOHYD 37 37 N-LINKED (GLCNAC. .) (POTENTIAL).
CC FT CARBOHYD 37 37 N-LINKED (GLCNAC. .) (POTENTIAL).
CC FT CARBOHYD 46 48 N-LINKED (GLCNAC. .) (POTENTIAL).
CC FT LIPID 56 56 GPI-ANCHOR (POTENTIAL).
CC FT SEQUENCE 76 AA; 7862 MW; 42846E7DEC3D9D58 CRC64;
CC
CC Query Match 10.6%; Score 45; DB 1; Length 76;
CC Best Local Similarity 24.4%; Pred. No. 1.7e+02;
CC Matches 19; Conservative 10; Mismatches 15; Indels 34; Gaps 4;
CC
CC QY 6 CSQSISPMRSISNVLAMDFGSKQS-RVIENPTAELSAVEEGLAWRKGCGLRGTHG 64
CC 1: 1:1 111:1 111:1 111:1 111:1 111:1
CC 26 CNOTSVAP-----FSGNOSIAAAPTNTATT-----RSGC----- 55
CC ([1])
CC
CC DB 65 SPTASSQSATNMIAHRS 82
CC 1: 1:1 111:1 111:1 111:1 111:1 111:1
CC 56 ---SSLQSTAGLLALSLS 70
CC ([1])
CC
CC RESULT 5
CC WT2_MALDO STANDARD; PRT; 79 AA.
CC AC 024058;
CC 15-JUL-1998 (Rel. 36, Created)
CC 15-JUL-1998 (Rel. 36, Last sequence update)
CC 15-JUN-2002 (Rel. 41, Last annotation update)
CC Metallothionein-like protein type 2.
CC MTL
CC OS Malus domestica (Apple) (Malus sylvestris).
CC OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
CC OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
CC OC eurosids I; Rosales; Rosaceae; Maloideae; Malus.
CC NCBI_Taxid:3750;
CC ([1])
CC SEQUENCE FROM N.A.
CC TISSUE=Fruit cortical tissue;
CC REID S.J., Ross G.S.;
CC "Up-regulation of two cDNA clones encoding metallothionein-like
CC proteins in apple fruit during cool storage.";
CC Physiol. Plantarum 100:183-189(1997).
CC -!- FUNCTION: METALLOTHIONEINS HAVE A HIGH CONTENT OF CYSTEINE
CC -!- RESIDUES THAT BIND VARIOUS HEAVY METALS.
CC -!- SIMILARITY: BELONGS TO THE METALLOTHIONEIN SUPERFAMILY; FAMILY 15.

```



OS Methanococcus jannaschii.  
 OC Archaea; Euryarchaeota; Methanococci; Methanococcales;  
 OC Methanocaldococcaceae; Methanocaldococcus.  
 OX NCBI\_TaxID=2190;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;  
 RX MEDLINE=96337999; PubMed=8688087;  
 RA Sult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,  
 RA Sutcliffe G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,  
 RA Kierulff A.R., Dougherty B.A., Tomb J.F., Adams M.D., Reich C.I.,  
 RA Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Glodek A.,  
 RA Uterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,  
 RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,  
 RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.,  
 RT "Complete genome sequence of the methanogenic archaeon, Methanococcus  
 jannaschii.";  
 RL Science 273:1058-1073(1996).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: U67505; AAB98558.1;  
 DR TIGR: MJ0567;  
 DR KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 82 AA; 8766 MW; 3F3810EEFC9F81CE CRC64;  
 Query Match 10.2%; Score 43; DB 1; Length 82;  
 Best Local Similarity 25.9%; Pred. No. 3.1e+02;  
 Matches 15; Conservative 10; Mismatches 15; Indels 18; Gaps 3;  
 QY 4 GSSSSQSPMSRISLVMDFSGOKSRVIENTPEALSVAVEGLAWR 53  
 DB 20 AGCGAM-----ORLVSGINGISKLKVRNQNGPVISTKGSNIAIGRLANK 67  
 RESULT 9  
 ID HFO2\_METFO STANDARD; PRT; 67 AA.  
 AC P48783;  
 DT 01-FEB-1996 (Rel. 33, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Archaeal histone A2.  
 GN HFOA2.  
 OS Methanobacterium formicicum.  
 OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;  
 OC Methanobacteriaceae; Methanobacterium.  
 OX NCBI\_TaxID=2162;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=JF-1;  
 RX MEDLINE=95138058; PubMed=7836329;  
 RA Darcy T.J., Sandman K.M., Reeve J.N.;  
 RT "Methanobacterium formicicum, a mesophilic methanogen, contains three  
 RT HFO histones.";  
 RL J. Bacteriol. 177:858-860(1995).  
 RN (2)  
 RP PARTIAL SEQUENCE.  
 RX MEDLINE=95138058; PubMed=7836329;  
 RA Sandman K.M., Grayling R.A., Reeve J.N.;  
 RA Unpublished results, cited by:  
 RL Darcy T.J., Sandman K.M., Reeve J.N.;  
 RL J. Bacteriol. 177:858-860(1995).  
 CC -!- FUNCTION: BINDS AND COMPACT DNA (95 TO 150 BASE PAIRS) TO FORM  
 CC NUCLEOSOME-LIKE STRUCTURES THAT CONTAIN POSITIVE DNA SUPERCOILS.  
 CC -!- SUBUNIT: HOMODIMER OR HETERODIMER (POTENTIAL).  
 CC -----

CC -!- SIMILARITY: BELONGS TO THE HMF FAMILY OF ARCHAEAL HISTONES.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: U12031; AAA67722.1;  
 DR HSP: P48781; 1567;  
 DR InterPro: IPR003958; CBFA\_NFYB\_domain.  
 DR InterPro: IPR004822; Histone\_core.  
 DR Pfam: PF00808; CBFA\_NFYB\_HMF; 1.  
 DR DNA-binding; Multigene family.  
 KW INIT\_MET 0  
 FT INIT\_MET 0  
 SQ SEQUENCE 67 AA; 7064 MW; 0AAFCAC535BF2E10 CRC64;  
 Query Match 10.0%; Score 42.5; DB 1; Length 67;  
 Best Local Similarity 25.8%; Pred. No. 2.8e+02;  
 Matches 16; Conservative 12; Mismatches 25; Indels 9; Gaps 2;  
 QY 11 ISPMRSISLVMDFSGOKSRVIENTPEALSVAVEGLAWRKGCCLRLGTH-QSPTAS 69  
 DB 5 IAPVGRILKNA-----GAQRISDDAKEALAKALENGEELAKKAVELAKHAGRKTVK 56  
 QY 70 SQ 71  
 DB 57 AE 58  
 RESULT 10  
 ID RS19\_CHLIR STANDARD; PRT; 88 AA.  
 AC O84529; Q9PJL8;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE 30S ribosomal protein S19.  
 GN RPSS OR RS19 OR CT524 OR TC0811.  
 OS Chlamydia trachomatis, and  
 OS Chlamydia muridarum.  
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
 OX NCBI\_TaxID=813, 83560;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=C. trachomatis; STRAIN=D/UW-3/Cx;  
 RX MEDLINE=9900809; PubMed=9784136;  
 RA Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,  
 RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,  
 RA Davis R.W.;  
 RT "Genome sequence of an obligate intracellular pathogen of humans:  
 RT Chlamydia trachomatis.";  
 RL Science 282:754-759(1998).  
 RN (2)  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=C. muridarum; STRAIN=MoPn / Nigg;  
 RX MEDLINE=20150255; PubMed=10684935;  
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,  
 RA White O., Hickey E.K., Peterson J., Uterback T., Berry K., Bass S.,  
 RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,  
 RA Gwin M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,  
 RA Eisen J., Fraser C.M.;  
 RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia  
 RT pneumoniae AR39.";  
 RL Nucleic Acids Res. 28:1397-1406(2000).  
 CC -!- FUNCTION: PROTEIN S19 FORMS A COMPLEX WITH S13 THAT BINDS STRONGLY  
 CC TO THE 16S RIBOSOMAL RNA (BY SIMILARITY).  
 CC -!- SIMILARITY: BELONGS TO THE S19P FAMILY OF RIBOSOMAL PROTEINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -





```

Query Match          9.9%; Score 42; DB 1; Length 78;
Best Local Similarity 27.3%; Pred. No. 3.8e+02;
Matches 15; Conservative 11; Mismatches 21; Indels 8; Gaps 2;

QY 8 SOSISPMRSISENVAMDFSGOKSRVIENPTAELSAVVE---EGLAWRKKGCLR 59
    ||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 23 SQTLPQPMINAMLQQLSGRAVRIQEPDGLSGCKSCPEG-----KACLR 72

RESULT 13
TMOB_PSEME
ID TMOB_PSEME STANDARD; PRT; 83 AA.
AC Q00457;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Toluene-4-monooxygenase system protein B (BC 1.14.13.-).
GN TMOB.
OS Pseudomonas mendocina.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=300;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-27.
RC STRAIN=KRI;
RX MEDLINE=91358306; PubMed=1885512;
RA Yen K.-M., Karl M.R., Blatt L.M., Simon M.J., Winter R.B.,
RA Fausset P.R., Lu H.S., Harcourt A.A., Chen K.K.;
RT "Cloning and characterization of a Pseudomonas mendocina KRI gene
RT cluster encoding toluene-4-monooxygenase.";
RL J. Bacteriol. 173:5315-5327(1991).
CC -!- FUNCTION: HYDROXYLATES TOLUENE TO FORM P-CRESOL.
CC -!- COFACTOR: FAD; REQUIRES FE(+2) FOR ACTIVITY.
CC -!- PATHWAY: Toluene degradation; first step.
CC -!- SUBUNIT: THE MULTICOMPONENT ENZYME TOLUENE-4-MONOOXYGENASE
CC IS FORMED BY THE TMOA, TMOB, TMOG, TMOH, TMOI, TMOJ AND TMOF
CC POLYPEPTIDES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M65106; AAA36000.1;
KW Aromatic hydrocarbons catabolism; Oxidoreductase; Flavoprotein;
KW Monooxygenase; FAD; iron.
FT INIT_MET 0
SQ SEQUENCE 83 AA; 9457 MW; 4729FEF73F266F44 CRC64;

Query Match          9.9%; Score 42; DB 1; Length 83;
Best Local Similarity 26.5%; Pred. No. 4.1e+02;
Matches 13; Conservative 9; Mismatches 19; Indels 8; Gaps 2;

QY 6 CSSQISIP-----MRSISENVAMDFSGOKSRVIENPTAELSAVVE 48
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 37 CVNRVAPREGVMVRKRRSTELFPRDPTIAESGL--NPTEVIDVVEE 83

RESULT 14
HFOB_METFO
ID HFOB_METFO STANDARD; PRT; 67 AA.
AC P48784;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Archaeal histone B.
GN HFOB.
OS Methanobacterium formicicum.
OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;

```

```

Methanobacteriaceae; Methanobacterium.
NCBI_TaxID=2162;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JF-1;
RX MEDLINE=95138058; PubMed=7836329;
RA Darcy T.J., Sandman K.M., Reeve J.N.;
RT "Methanobacterium formicicum, a mesophilic methanogen, contains three
RT HFO histones.";
RL J. Bacteriol. 177:858-860(1995).
CC -!- FUNCTION: BINDS AND COMPACT DNA (95 TO 150 BASE PAIRS) TO FORM
CC NUCLEOSOME-LIKE STRUCTURES THAT CONTAIN POSITIVE DNA SUPERCOILS.
CC -!- SUBUNIT: HOMODIMER OR HETERODIMER (POTENTIAL).
CC -!- SIMILARITY: BELONGS TO THE HMF FAMILY OF ARCHAEL HISTONES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U12929; AAA67720.1;
DR HSSP; P48781; 1B67.
DR InterPro; IPR003958; CBFA_NFYB_domain.
DR InterPro; IPR004822; Histone_core.
DR Pfam; PF00808; CBFD_NFYB_HMF; 1.
KW DNA-binding; Multigene family.
SQ SEQUENCE 67 AA; 7149 MW; 1132F83ACAD88445 CRC64;

Query Match          9.8%; Score 41.5; DB 1; Length 67;
Best Local Similarity 26.9%; Pred. No. 3.6e+02;
Matches 21; Conservative 14; Mismatches 20; Indels 23; Gaps 5;

QY 11 ISPMRSISENVAMDFSGOKSRVIENPTAELSAVVEGLAWRKKG-----CLRGLTH- 63
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 5 IAPGILKNA-----GAERVSDAREALAKALEE-----KGTIATEAVKLAKHA 50

QY 64 GSPTASSOSSATNMAIHR 81
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 51 GRKTV--RASDVELAVKR 66

RESULT 15
RS19_MYCGA
ID RS19_MYCGA STANDARD; PRT; 87 AA.
AC O52336;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE 30S ribosomal protein S19.
GN RPSS OR RPS19.
OS Mycoplasma gallisepticum.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2096;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A5969Var.B;
RA Skamrov A., Feoktistova E., Goldman M., Bebealashvili R.;
RT "Cloning and sequencing of Mycoplasma gallisepticum chromosome region
RT containing operons S10 and rin235.";
RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: PROTEIN S19 FORMS A COMPLEX WITH S13 THAT BINDS STRONGLY
CC TO THE 16S RIBOSOMAL RNA (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE S19P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/

```

DR PROSITE; PS00190; CYTOCHROME C; 1.

FT	BINDING	17	HEME (COVALENT)
FT	BINDING	17	HEME (COVALENT)
FT	METAL	18	IRON (HEME AXIAL LIGAND)
FT	METAL	60	IRON (HEME AXIAL LIGAND)
FT	CONFLICT	35	K -> L (IN REF. 1)
FT	CONFLICT	79	YASLPADGGS -> GASXP (IN REF. 1)
SQ	SEQUENCE	88 AA;	CERFEFEBG9IAB2E0 CRC64;
		Query Match	Score 41.5; DB:1; Length 88;
		Best local similarity	22.8%; Pred. No. 5e+02

	MALCHES	19;	CONSERVATIVE	13;	MISMAKUCHES	23;	INDICLS	13;	GAPS	2
QY	2	GRSCSSQSITSPMSISENSLVAMDFGQKSRSRVIENTPALSVAVEGLAWRKKGCIRLG	61							
	I::I:			:		:	I::I:	: :~	.	
D8	19	GONGISOVPYP-----NLAQRE-----OYLVAALKAYKAGORGGOPAV	59							
QY	62	THGSPTASSQSASATNMAIH	80							
	I		:	I	:					
D8	60	MOGOATALSDADIANLAAY	78							

09477R1; 09JF78; 19-MAY-2000 (rel. 39, Created)  
15-JUN-2002 (rel. 41, Last sequence update)  
30-JUN-2002 (rel. 41, Last annotation update)  
30S: ribosomal protein S19;  
RPS8 OR RS19 OR CPN0643 OR CPN104  
Chlamydia pneumoniae (Chlamydia pneumoniae).  
Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
NCBI\_TaxId=83538;  
[1]  
SEQUENCE FROM N.A.  
STRAIN=CM1029;  
MEDLINE=99206606; PubMed=10192388;  
Kalmán S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,  
Lingler L., Grimwood J., Davis R.W., Stephens R.S.;  
"Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";  
Nat. Genet. 21:385-389(1999).  
[2]  
SEQUENCE FROM N.A.  
STRAIN=AR339;  
MEDLINE=20150255; PubMed=10684935;  
Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,  
White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,  
Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,  
Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,  
Eisen J., Fraser C.M.;  
"Genome sequences of Chlamydia trachomatis MoPn and Chlamydia  
pneumoniae AR39.";  
Nucleic Acids Res. 28:1397-1406(2000).  
[3]  
SEQUENCE FROM N.A.  
STRAIN=J138;  
MEDLINE=20330349; PubMed=10871362;  
Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,  
Shiba T., Ishik K., Hattori M., Kuhara S., Nakazawa T.;  
"Comparison of whole genome sequences of Chlamydia pneumoniae J138  
from Japan and CM1029 from USA.";  
Nucleic Acids Res. 28:2311-2314(2000).  
-!- FUNCTION: PROTEIN S19 FORMS A COMPLEX WITH S13 THAT BINDS STRONGLY  
TO THE 16S RIBOSOMAL RNA (BY SIMILARITY).  
-!- SIMILARITY: BELONGS TO THE S19P FAMILY OF RIBOSOMAL PROTEINS.  
-----  
THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
purposes is prohibited.

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; AE001647; AAD18782.1; -

DR EMBL; AE002173; AAF73629.1; -

DR EMBL; AF002547; BAA98850.1; -

DR HSP; P80381; 10KF.

DR TIGR; CP0104; -

DR InterPro; IPR002222; Ribosomal\_S19.

DR Pfam; PF00203; Ribosomal\_S19; 1.

DR ProDom; PD001012; Ribosomal\_S19; 1.

DR TIGRFAMs; TIGR01050; rpsBact; 1.

DR PROSITE; PS00323; RIBOSOMAL\_S19; 1.

DR Ribosomal protein; rRNA-binding; C (IN REF.).

FT CONFLICT 2

FT SEQUENCE 88 AA; 10201 MW; EBL98E3B4BC2588 CRC64;

SQ

Query Match 9.8%; Score 41.5; DB 1; Length 88;

Best Local Similarity 26.7%; Pred. No. 5e+02;

Matches 12; Conservative 10; Mismatches 22; Indels 1; Gaps 1;

OY 12 SPMSISNSLVAMDFSGQSKSVIENPTALSVAVEGLAWRKKG 56

Db

29 TPKTWSSRMITPEMIGHTFEV-HNGKFLTVFVSETMGHKL 72

RESULT 18

RPN\_SULSO

ID RPN\_SULSO STANDARD; PRT; 66 AA.

AC Q980Z8;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE DNA-directed RNA polymerase subunit N (EC 2.7.7.6).

GN RPN OR SS05140.

OS Sulfolobus solfataricus.

OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;

OC Sulfolobus.

OX NCBI\_TaxID=2287;

RN

SEQUENCE FROM N.A.

RC STRAIN-ATCC 35092 / DSM 1617 / P2;

RX MEDLINE=21332296; PubMed=11427726;

RA She O., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,

RA Aweez M.J., Chan-Weiher C.C.-Y., Clausen I.G., Curtis B.A.,

RA De Moors A., Erauso G., Fletcher C., Gordon P.M.K.,

RA Heikamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,

RA Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,

RA Charlebois R.L., Doolittle W.F., Duquet M., Gaasterland T.,

RA Garrett R.A., Ragan M.A., Sensen C.W., Van der Oost J.

RL "The complete genome of the crenarchaeon Sulfolobus solfataricus P2."

RT Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).

CC -!- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS SUBSTRATES.

CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate + (RNA)(N).

CC -!- SUBUNIT: THE S. ACIDOCALDIARUS RNAP IS COMPOSED OF 13 SUBUNITS.

CC -!- SIMILARITY: BELONGS TO THE ARCHAEABACTERIA RPN / EUKARYOTIC RPB10 RNA POLYMERASE SUBUNIT FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; AE006647; AAK40429.1; -

DR InterPro; IPR000268; RNA\_POL\_N.

DR Pfam; PF01194; RNA\_POL\_N; 1.

DR PROSITE; PS01112; RNA\_POL\_N\_8KD; 1.

KW Transferase; DNA-directed RNA polymerase; Transcription; Zinc;

FT METAL 7 7 ZINC (BY SIMILARITY).

FT METAL 10 10 ZINC (BY SIMILARITY).

FT METAL 44 44 ZINC (BY SIMILARITY).

FT METAL 45 45 ZINC (BY SIMILARITY).

SQ SEQUENCE 66 AA; 7591 MW; C6774B541A1CFA13 CRC64;

Query Match 9.7%; Score 41; DB 1; Length 66;

Best Local Similarity 25.7%; Pred. No. 4.1e+02;

Matches 19; Conservative 12; Mismatches 29; Indels 3;

OY 13 PMRSISNSLVAMDFSGQSKSVI-ENPTALSVAVEGLAWRKKGCLRLGTHGSPPASS 70

Db

4 PIRCFCTGSLIADKQSFITRVNAGENPGKVL-----DDLGVKRYCCRRM-----LLS 51

OY 71 QSSATNMAIHRSP 84

Db

52 HVDIINEVIHYTRP 65

RESULT 19

CC3\_CARCN

ID CC3\_CARCN STANDARD; PRT; 43 AA.

AC P32956;

DT 01-OCT-1993 (Rel. 27, Created)

DT 01-OCT-1993 (Rel. 27, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Cysteine proteinase III (EC 3.4.22.-) (CC-III) (Fragment).

OS Carica candamarcensis.

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC eurosids II; Brassicales; Caricaceae; Carica.

OX NCBI\_TaxID=29731;

RN

SEQUENCE.

RC TISSUE=Latex;

RX MEDLINE=94030669; PubMed=8216902;

RA Walravens V., Jaziri M., van Beeumen J., Schneek A.G.,

RA Kleinschmidt T., Looze Y.

RT "Isolation and preliminary characterization of the cysteine-

RL Biol. Chem. Hoppe-Seyler 374:501-506(1993).

CC -!- PTM: GLYCOSYLATED

CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C1.

DR HSP; P14080; ITAL.

DR MEROPS; COL.020; -

DR InterPro; IPR000668; Peptidase\_C1.

DR InterPro; IPR000169; SHprot\_acsite.

DR Pfam; PF00112; Peptidase\_C1; 1.

DR ProDom; PD000158; Peptidase\_C1; 1.

DR PROSITE; PS00639; THIOLEPROTEASE\_HIS; PARTIAL.

DR PROSITE; PS00640; THIOLEPROTEASE\_ASN; PARTIAL.

DR PROSITE; PS00139; THIOLEPROTEASE\_CYS; 1.

KW Hydrolase; Thiol protease; Glycoprotein.

FT ACT\_SITE 25 25 BY SIMILARITY.

FT NON\_TER 43 43

SQ SEQUENCE 43 AA; 4636 MW; F4C5D2881886291 CRC64;

Query Match 9.6%; Score 40.5; DB 1; Length 43;

Best Local Similarity 32.5%; Pred. No. 2.8e+02;

Matches 13; Conservative 5; Mismatches 15; Indels 7; Gaps 2;

OY 48 EGLAWRKKGCI-----RLGTHGSPASSQSAT---NMAIH 80

Db

3 ESIDRRKKGAVTPVKNQSGCGSCWAFSTIATVEGINKIVH 42

RESULT 20

GBG5\_HUMAN

ID GBG5\_HUMAN STANDARD; PRT; 68 AA.

AC P30670; Q61015;

DT 01-APR-1993 (Rel. 25, Created)

DT 01-APR-1993 (Rel. 25, Last sequence update)  
 DE 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Guanine nucleotide-binding protein G(I)/G(S)/G(O) gamma-5 subunit.  
 GN GNG5 OR GNGT5.  
 OS Homo sapiens (Human),  
 OS Mus musculus (Mouse),  
 OS Rattus norvegicus (Rat), and  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606, 10090, 10116, 9913,  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=Human;  
 RX MEDLINE=99009227; PubMed=9790912;  
 RA Yu B., Aronson N.N. Jr.;  
 RL "Structure of human G protein Ggamma5 gene GNG5.";  
 RL Biochem. Biophys. Res. Commun. 251:88-94(1998).  
 RN (2)  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=Human; TISSUE=Blood;  
 RX MEDLINE=98318631; PubMed=9653160;  
 RA Mao M., Fu G., Wu J.-S., Zhang Q.-H., Zhou J., Kan L.-X., Huang Q.-H.,  
 RA He K.-L., Gu B.-W., Han Z.-G., Shen Y., Gu J., Yu Y.-P., Xu S.-H.,  
 RA Wang Y.-X., Chen S.-J., Chen Z.  
 RT "Identification of genes expressed in human CD34(+) hematopoietic  
 RT stem/progenitor cells by expressed sequence tags and efficient full-  
 RT length cDNA cloning.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 95:8175-8180(1998).  
 RN (3)  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=Human; TISSUE=Brain;  
 RX Strausberg R.;  
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
 RN (4)  
 RP SEQUENCE FROM N.A.  
 RC SPECIES=Bovine, and Rat; TISSUE=Liver;  
 RX MEDLINE=92195304; PubMed=1549114;  
 RA Fisher K.J., Aronson N.N. Jr.;  
 RT "Characterization of the cDNA and genomic sequence of a G protein  
 RT gamma subunit (gamma 5).";  
 RL Mol. Cell. Biol. 12:1585-1591(1992).  
 RN (5)  
 RP SEQUENCE.  
 RC SPECIES=Bovine; TISSUE=Spleen;  
 RX MEDLINE=93356792; PubMed=8352779;  
 RA Morishita R., Masuda K., Niwa M., Kato K., Asano T.;  
 RT "Identification of three forms of the gamma subunit of G proteins  
 RT isolated from bovine spleen.";  
 RL Biochem. Biophys. Res. Commun. 194:1221-1227(1993).  
 RN (6)  
 RP SEQUENCE OF 8-53 FROM N.A.  
 RC SPECIES=Mouse; STRAIN=CF-1 / Harlan;  
 RX MEDLINE=97011591; PubMed=8858601;  
 RA Williams C.J., Schultz R.M., Kopf G.S.;  
 RT "G protein gene expression during mouse oocyte growth and maturation,  
 RT and preimplantation embryo development.";  
 RL Mol. Reprod. Dev. 44:315-323(1996).  
 CC -1- FUNCTION: GUANINE NUCLEOTIDE-BINDING PROTEINS (G PROTEINS) ARE  
 CC INVOLVED AS A MODULATOR OR TRANSDUCER IN VARIOUS TRANSMEMBRANE  
 CC SIGNALING SYSTEMS. THE BETA AND GAMMA CHAINS ARE REQUIRED FOR THE  
 CC GTPASE ACTIVITY, FOR REPLACEMENT OF GDP BY GTP, AND FOR G PROTEIN-  
 CC EFFECTOR INTERACTION.  
 CC -1- SUBUNIT: G PROTEINS ARE COMPOSED OF 3 UNITS (ALPHA, BETA & GAMMA).  
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN A VARIETY OF TISSUES.  
 CC -1- MISCELLANEOUS: THERE ARE AT LEAST SEVEN DIFFERENT GAMMA SUBUNITS  
 CC WITH POSSIBLE TISSUE-SPECIFIC DISTRIBUTION.  
 CC -1- SIMILARITY: BELONGS TO THE G PROTEIN GAMMA FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: L22475; AAA03621.1;  
 CC Genew: HGNC:959; BAX.

CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: AF085709; AAC72203.1;  
 CC EMBL: AF085709; AAC72203.1; JOINED.  
 CC EMBL: AC039355; AAC39869.1;  
 CC EMBL: BC003563; AAA03563.1;  
 CC EMBL: M57779; AAA30535.1;  
 CC EMBL: M57780; AAA41188.1;  
 CC EMBL: U38498; AAB01729.1;  
 CC PIR: JN0701; JN0701.  
 CC PIR: B42243; B42243.  
 CC Genew: HGNC:4408; GNG5.  
 CC MIM: 600874;  
 CC MIM: MGI:109164; Gng5.  
 CC InterPro: IPR001770; G-gamma.  
 CC Pfam: PF00631; G-gamma; 1.  
 CC PRINTS: PR00321; GPROTEIN.  
 CC ProDom: PD003783; G-gamma; 1.  
 CC SMART: SM00224; GGL; 1.  
 CC PROSITE: PS00058; G-PROTEIN\_GAMMA; 1.  
 CC Transducer; Prenylation; Lipoprotein; Lipoprotein; Multigene family.  
 KW TRANSFER; GERANYL-GERANYL (BY SIMILARITY).  
 FT LIPID 65 65  
 FT PROPEP 66 68  
 FT SEQUENCE 68 AA; 7318 MW; 9AF7A16558863602 CRC64;  
 SQ  
 Query Match 9.6%; Score 40.5; DB 1; Length 68;  
 Best Local Similarity 36.4%; Pred No. 4.8e+02;  
 Matches 12; Conservative 5; Mismatches 15; Indels 1; Gaps 1;  
 Qy 44 VAVEEGLAWRKKGCLRLGTHCSPTASSQSSATN 76  
 Db 26 VKVQAADLKQFCLQNAQH-DPLLTVGVSSSTN 57  
 RESULT 21  
 ID BAXC\_HUMAN STANDARD; PRT; 41 AA.  
 AC Q07815;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE BAX protein, cytoplasmic isoform gamma.  
 GN BAX.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=B-cell.  
 RX MEDLINE=93364978; PubMed=8358790;  
 RA Olveit Z.N., Millman C.L., Korsmeyer S.J.;  
 RT "Bcl-2 heterodimerizes in vivo with a conserved homolog, Bax, that  
 RT accelerates programmed cell death.";  
 RL Cell 74:609-619(1993).  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -1- ALTERNATIVE PRODUCTS: THE MEMBRANE ISOFORM ALPHA AND THE THREE  
 CC CYTOPLASMIC ISOFORMS, BETA, GAMMA AND DELTA ARE GENERATED BY  
 CC ALTERNATIVE SPLICING.  
 CC -1- SIMILARITY: BELONGS TO THE BCL-2 FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: L22475; AAA03621.1;  
 CC Genew: HGNC:959; BAX.

```

DR MM: 600040;
KW Apoptosis; Alternative splicing.
SQ SEQUENCE 41 AA; 4678 MW; D94639AABB927859 CRC64;

Query Match
Best Local Similarity 9.5%; Score 40; DB 1; Length 41;
Matches 13; Conservative 5; Mismatches 12; Indels 12; Gaps 2;

QY 25 MDFSQKSRVTEPTALSVAVEEG-----LAWKKGKGLRL 60
||| :|||
DB 1 MDGSG-----EQPRGVSSRIEQEGWGGRHPSNPWTCLRM 36

RESULT 22
HS2M-LYCES STANDARD; PRT; 56 AA.
AC P81161;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Heat shock 22 kDa protein, mitochondrial (fragments).
OS Lycopersicon esculentum (Tomato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
OX NCBI_TaxID=4081;
RN [1]
RP SEQUENCE.
RC STRAIN=CV. SWEET;
RX MEDLINE=98345975; PubMed=9680997;
RA Banzet N., Richard C., Deveaux Y., Kazmaier M., Gagnon J.,
RA Triantaphyllides C.;
RT "Accumulation of small heat shock proteins, including mitochondrial
RT HSP22, induced by oxidative stress and adaptive response in tomato
RT cells";
RL Plant J. 13:519-527(1998).
CC -!- FUNCTION: MAY PLAY A PROTECTIVE ROLE AGAINST OXIDATIVE STRESS.
CC -!- SUBCELLULAR LOCATION: Mitochondrial.
CC -!- INDUCTION: BY HEAT SHOCK, AND UNDER OTHER CONDITIONS OF STRESS,
CC SUCH AS INCREASED SALT CONCENTRATION AND STARVATION.
CC -!- SIMILARITY: BELONGS TO THE SMALL HEAT SHOCK PROTEIN (HSP20)
CC FAMILY.
DR InterPro: IPR002066; HSP20.
DR PROSITE: PS01031; HSP20; PARTIAL.
KW Heat shock; Mitochondrion.
FT NON_CONS 14 15
FT UNSURE 15 15
FT NON_CONS 35 36
FT NON_CONS 36 36
FT UNSURE 36 36
FT NON_TER 56 56
SQ SEQUENCE 56 AA; 6446 MW; 2AB9F927C7720076 CRC64;

Query Match
Best Local Similarity 9.5%; Score 40; DB 1; Length 56;
Matches 9; Conservative 4; Mismatches 10; Indels 0; Gaps 0;

QY 38 PTEALSVAVEEGLAWKKGKGLRL 60
||| :|||
DB 21 PVENRVALEENTLIMKNGVLKV 43

RESULT 23
RPON_THEAC STANDARD; PRT; 72 AA.
AC Q9HL09;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE DNA-directed RNA polymerase subunit N (EC 2.7.7.6).
GN RPON OR TA0431.
OS Thermoplasma acidophilum.
OC Archaea; Euryarchaeota; Thermoplasma; Thermoplasmatales;
OC Thermoplasmataceae; Thermoplasma.

```

```

OX NCBI_TaxID=2303;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM 1728;
RX MEDLINE=20479972; PubMed=11029001;
RA Ruepp A., Graml W., Santos-Martinez M.-L., Koretke K.K., Volker C.,
RA Neues H.-W., Frishman D., Stocker S., Lupas A.N., Baumeister W.;
RT "The genome sequence of the thermoacidophilic scavenger Thermoplasma
RT acidophilum";
RL Nature 407:308-313(2000).
CC -!- FUNCTION: DNA-dependent RNA polymerase catalyzes the transcription
CC of DNA into RNA using the four ribonucleoside triphosphates as
CC substrates.
CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC [RNA](N).
CC -!- SIMILARITY: BELONGS TO THE ARCHAEABACTERIA RPON / EUKARYOTIC RPB10.
CC RNA POLYMERASE SUBUNIT FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed, usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
DR EMBL: AL445064; CAC11573.1;
DR HSP: O26147; 1EF4
DR IPR: IPR000268; RNA_pol_N.
DR Pfam: PF01194; RNA_pol_N; 1.
DR PROSITE: PS01112; RNA_POL_N-BKD; 1.
KW Transferrase; DNA-directed RNA polymerase; Transcription; Zinc;
KW Metal-binding; Complete proteome.
FT METAL 7 7 ZINC (BY SIMILARITY).
FT METAL 10 10 ZINC (BY SIMILARITY).
FT METAL 53 53 ZINC (BY SIMILARITY).
FT METAL 54 54 ZINC (BY SIMILARITY).
SQ SEQUENCE 72 AA; 8368 MW; 792AEDA20E5447E2 CRC64;

Query Match
Best Local Similarity 9.5%; Score 40; DB 1; Length 72;
Matches 12; Conservative 5; Mismatches 21; Indels 0; Gaps 0;

QY 11 ISPMRSISENSLVAMDFSGQKSRVIENTEALSVAVEE 48
||| :|||
DB 2 IIPVRCFSGRVIASDYGRYIKRVNEIKAEGRDPSPPE 39

RESULT 24
RPON_THEVO STANDARD; PRT; 72 AA.
AC Q979K0;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE DNA-directed RNA polymerase subunit N (EC 2.7.7.6).
GN RPON OR TV1161 OR TVG1188103.
OS Thermoplasma volcanium.
OC Archaea; Euryarchaeota; Thermoplasma; Thermoplasmatales;
OC Thermoplasmataceae; Thermoplasma.
OX NCBI_TaxID=50339;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GSS1 / DSM 4299 / JCM 9571;
RX MEDLINE=20570466; PubMed=11121031;
RA Kawashima T., Amano N., Koike H., Makino S.-I., Higuchi S.,
RA Kawashima-Ohya Y., Watanabe K., Yamazaki M., Kanehori K., Kawamoto T.,
RA Nunoshiba T., Yamamoto Y., Aramaki H., Makino K., Suzuki M.;
RT "Archaeal adaptation to higher temperatures revealed by genomic
RT sequence of Thermoplasma volcanium";
RL Proc. Natl. Acad. Sci. U.S.A. 97:14257-14262(2000).
CC -!- FUNCTION: DNA-dependent RNA polymerase catalyzes the transcription
CC of DNA into RNA using the four ribonucleoside triphosphates as

```

CC substrates.  
CC -I- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +  
CC (RNA)(N).  
CC -I- SIMILARITY: BELONGS TO THE ARCHAEABACTERIA RPON / EUKARYOTIC RPBL0  
CC RNA POLYMERASE SUBUNIT FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
CC EMBL: AP000995; BAB60303.1;  
CC InterPro: IPR000268; RNA\_POL\_N.  
CC Pfam: PF01194; RNA\_POL\_N.1.  
CC PROSITE: PS01112; RNA\_POL\_N.8KD.1.  
CC Transferrase: DNA-directed RNA polymerase; Transcription; Zinc;  
CC Metal-binding; Complete proteome.  
CC FT METAL 7 7 ZINC (BY SIMILARITY).  
CC FT METAL 10 10 ZINC (BY SIMILARITY).  
CC FT METAL 53 53 ZINC (BY SIMILARITY).  
CC FT METAL 54 54 ZINC (BY SIMILARITY).  
CC SEQUENCE 72 AA; 8483 MW; 06AEC0AA7AC73CA6 CRC64;

Query Match 9.5%; Score 40; DB 1; Length 72;  
Best Local Similarity 28.9%; Pred. No. 5.9e+02;  
Matches 11; Conservative 6; Mismatches 21; Indels 0; Gaps 0;

QY 11 ISPMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEE 48  
| | | | : | | : | | : | |  
DB 2 IIPVRCGGRVIASDYGRLRLRINEIRSEGREPTAE 39

RESULT 25  
V187\_BPT7  
ID V187\_BPT7 STANDARD; PRT; 83 AA.  
AC P03788;  
DT 21-JUL-1986 (Rel. 01, Created).  
DT 21-JUL-1986 (Rel. 01, Last sequence update).  
DT 01-MAR-1989 (Rel. 10, Last annotation update).  
DE Gene 18.7 protein.  
GN 18.7.  
OS Bacteriophage T7.  
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Podoviridae;  
OC T7-like viruses.  
OX NCBI\_TaxID=10760;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=83241725; PubMed=6864790;  
RA Dunn J.J., Studier F.W.;  
RT "Complete nucleotide sequence of bacteriophage T7 DNA and the  
RL J. Mol. Biol. 166:477-535(1983).  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: V01146; CA24439.1;  
DR PIR: A04413; W8BPG7.  
DR PIR: S42337; S42337.  
SQ SEQUENCE 83 AA; 9326 MW; B3DB055DE79A53C5 CRC64;

Query Match 9.3%; Score 39.5; DB 1; Length 83;  
Best Local Similarity 34.1%; Pred. No. 7.9e+02;  
Matches 15; Conservative 8; Mismatches 8; Indels 13; Gaps 3;

OY 4 SCSSQSSISPMRSISENSLVAMDFSGQKSRVIENPTEALSVAVEE 44  
| | | | : | | : | | : | |  
DB 47 SCASESKLP-----EPPMVSYD-----SSLMVEENLTTEMLNY 80

Search completed: March 28, 2003, 09:06:08  
Job time : 24.5197 secs

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 28, 2003, 09:02:45 ; Search time 33.7323 Seconds

(without alignments)  
239.394 Million cell updates/sec

File: us-09-936-697-6  
Perfect score: 423

Sequence: 1 QGRSCSSQSIIPMRISSEN.....SPTASSQSSATNMIAHRSQP 84

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 22614

Minimum DB seq length: 40

Maximum DB seq length: 90

Post-processing: Minimum Match 0%

Maximum Match 100%

-Listing first 1000 summaries

Database: April-73

1: Pir1 \*

2: Pir2 \*

3: Pir3 \*

4: Pir4 \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	54	12.8	80	2 T27603	hypothetical prote
2	51.5	12.2	76	2 E64324	DNA-directed RNA p
3	48	11.3	89	2 AG3246	conserved hypothet
4	47.5	11.2	90	2 F95310	hypothetical prote
5	46.5	11.0	62	2 E64510	hypothetical prote
6	46	10.9	77	2 S24471	gag polyprotein
7	46	10.9	87	2 D82861	hypothetical prote
8	45.5	10.8	65	2 A82567	hypothetical prote
9	45.5	10.8	73	2 T25763	hypothetical prote
10	45.5	10.8	81	2 JC5345	hypothetical prote
11	45	10.6	60	2 A42960	cdcl protein - Clo
12	45	10.6	76	2 T53107	ferredoxin 2[4Fe-4
13	45	10.6	75	2 T53107	CD24 precursor - r
14	45	10.6	78	2 H69420	hydrogenase expres
15	45	10.6	79	2 D81246	hypothetical prote
16	45	10.6	82	2 D69087	metallothionein-li
17	45	10.6	86	2 C41139	hydrogenase expres
18	44.5	10.5	64	2 AB2271	protein kinase 3 (
19	44.5	10.5	79	2 S75293	periplasmic mercur
20	44	10.4	51	2 D85807	hypothetical prote
21	44	10.4	78	2 B90959	hypothetical prote
22	44	10.4	78	2 A75411	probable phage tai
23	43.5	10.3	83	2 G90914	hypothetical prote
24	43.5	10.3	87	2 S60284	excisionase (impor
25	43.5	10.3	89	2 AH2653	BL2D protein - bar
26	43	10.2	82	1 G64370	hypothetical prote
27	42.5	10.0	52	2 D82682	conserved hypothet
28	42.5	10.0	64	2 AC2544	hypothetical prote
29	42.5	10.0	68	2 T00189	hypothetical prote

30	42.5	10.0	74	2 S72807	hypothetical prote
31	42.5	10.0	88	2 B71507	probable S19 ribos
32	42	9.9	43	2 S35580	cysteine proteinas
33	42	9.9	45	2 S36709	B7 protein - equin
34	42	9.9	53	2 C82776	hypothetical prote
35	42	9.9	64	2 C96993	hypothetical prote
36	42	9.9	73	2 F69062	molybdenum transp
37	42	9.9	78	2 D91160	hypothetical prote
38	42	9.9	78	2 C86006	hypothetical prote
39	42	9.9	78	2 E65136	hypothetical 8.7 k
40	42	9.9	79	2 A86517	hypothetical prote
41	42	9.9	79	2 B72106	hypothetical prote
42	42	9.9	81	2 D81565	hypothetical prote
43	42	9.9	86	2 AD1584	hypothetical prote
44	42	9.9	86	2 AE1230	hypothetical prote
45	42	9.9	89	2 A84086	hypothetical prote
46	41.5	9.8	72	2 A97136	zn-finger containi
47	41.5	9.8	77	2 B98247	hypothetical prote
48	41.5	9.8	80	2 S00858	T-cell receptor al
49	41.5	9.8	88	2 E72055	S19 ribosomal prot
50	41.5	9.8	88	2 H86570	S19 ribosomal prot
51	41	9.7	52	2 A34657	cathepsin E (EC 3.
52	41	9.7	60	2 I88870	interleukin 2 - we
53	41	9.7	62	2 I34512	interleukin 2 - mo
54	41	9.7	63	2 T03652	probable carbonate
55	41	9.7	66	2 F90146	DNA-directed RNA p
56	41	9.7	68	2 D53203	hypothetical prote
57	41	9.7	75	2 S24478	gag polyprotein -
58	41	9.7	86	2 E81144	hypothetical prote
59	40.5	9.6	68	2 B42243	GTP-binding regula
60	40.5	9.6	70	2 E70985	hypothetical prote
61	40.5	9.6	73	2 AG0172	proteoglycan core
62	40.5	9.6	77	2 AG3462	hypothetical prote
63	40	9.5	41	2 C47538	bcl-2-associated p
64	40	9.5	51	2 D81039	hypothetical prote
65	40	9.5	63	2 AC2553	hypothetical prote
66	40	9.5	64	2 S22044	gag polyprotein -
67	40	9.5	66	2 A43602	T-cell-stimulating
68	40	9.5	67	2 G97092	endoglucanase (tru
69	40	9.5	70	2 T02541	hypothetical prote
70	40	9.5	72	2 S24477	gag polyprotein -
71	40	9.5	73	2 S24472	gag polyprotein -
72	40	9.5	74	2 S24473	gag polyprotein -
73	40	9.5	75	2 S24475	late expression fa
74	40	9.5	75	2 S24474	voltage-dependent
75	40	9.5	76	2 S30403	putative small nuc
76	40	9.5	77	2 S31660	hypothetical prote
77	40	9.5	78	2 H90114	hypothetical prote
78	40	9.5	83	2 A82386	conserved hypothet
79	40	9.5	88	2 S77431	hypothetical prote
80	40	9.5	90	1 D69782	hypothetical prote
81	39.5	9.3	56	2 AE2413	4-oxalocrotonate t
82	39.5	9.3	61	2 F86696	probable transcrip
83	39.5	9.3	67	2 AF1487	repressor protein
84	39.5	9.3	67	2 AH1375	hypothetical prote
85	39.5	9.3	72	2 G97751	hypothetical prote
86	39.5	9.3	75	2 C81951	hypothetical prote
87	39.5	9.3	75	2 A70610	hypothetical prote
88	39.5	9.3	77	2 B95003	hypothetical prote
89	39.5	9.3	77	2 AD1945	hypothetical prote
90	39.5	9.3	81	2 F90454	hypothetical prote
91	39.5	9.3	83	1 W88PG7	gene 18.7 protein
92	39	9.2	45	1 C64901	ribosomal protein
93	39	9.2	45	1 D50889	30S ribosomal subu
94	39	9.2	45	2 PB5728	30S ribosomal subu
95	39	9.2	46	2 PC4152	toxin-co-regulated
96	39	9.2	62	2 T06654	hypothetical prote
97	39	9.2	64	2 A48411	Myf5 homolog - chi
98	39	9.2	69	2 T44123	hypothetical prote
99	39	9.2	72	2 AD3532	hypothetical prote
100	39	9.2	77	2 B83269	hypothetical prote
101	39	9.2	81	1 C70910	hypothetical prote
102	39	9.2	81	2 A97803	hypothetical prote



103	39	9.2	90	2	H72072	conserved hypothet	176	37	8.7	84	2	T35603	hypothetical prote
104	39	9.2	90	2	E86552	CT388 hypothetical	177	37	8.7	84	2	F83922	hypothetical prote
105	38.5	9.1	48	2	T35253	small hypothetical	178	37	8.7	84	2	AE3321	hypothetical prote
106	38.5	9.1	60	2	AC2981	hypothetical prote	179	37	8.7	87	2	F95373	hypothetical prote
107	38.5	9.1	63	2	T31143	hypothetical prote	180	37	8.7	88	2	S31030	gene 85 protein -
108	38.5	9.1	64	2	D81172	hypothetical prote	181	37	8.7	88	2	G82754	hypothetical prote
109	38.5	9.1	67	2	T42055	cold shock protein	182	37	8.7	89	2	F95314	conserved hypothet
110	38.5	9.1	67	2	C71854	hypothetical prote	183	36.5	8.6	46	2	G86705	hypothetical prote
111	38.5	9.1	67	2	A11126	probable transcrip	184	36.5	8.6	53	2	T35701	hypothetical prote
112	38.5	9.1	69	2	D69550	hypothetical prote	185	36.5	8.6	58	2	C83918	hypothetical prote
113	38.5	9.1	72	1	D75550	hypothetical prote	186	36.5	8.6	62	2	C71123	hypothetical prote
114	38.5	9.1	74	2	A25408	complement C5 - bo	187	36.5	8.6	71	2	B89990	hypothetical prote
115	38.5	9.1	76	2	B64660	hypothetical prote	188	36.5	8.6	78	1	E69549	conserved hypothet
116	38.5	9.1	76	2	A82122	hypothetical prote	189	36.5	8.6	79	2	B95045	hypothetical prote
117	38.5	9.1	78	1	EW8NYC	cinnamycin precurs	190	36.5	8.6	79	2	B84252	hypothetical prote
118	38.5	9.1	78	1	AC0997	conserved hypothet	191	36.5	8.6	79	2	T36418	hypothetical prote
119	38.5	9.1	82	2	G68162	SBH1 protein - yea	192	36.5	8.6	80	2	F90686	hypothetical prote
120	38.5	9.1	84	2	G96025	hypothetical prote	193	36.5	8.6	80	2	A85537	unknown (imported)
121	38.5	9.1	87	2	S73988	ribosomal protein	194	36.5	8.6	83	2	T79550	hypothetical camp
122	38.5	9.1	88	1	R3YMI9	ribosomal protein	195	36.5	8.6	83	2	A69940	305 ribosomal prot
123	38.5	9.1	89	2	AH1593	hypothetical prote	196	36.5	8.6	84	2	E69253	conserved hypothet
124	38	9.0	54	2	A31440	ovomucoid, third d	197	36.5	8.6	84	2	R3H523	hypothetical prote
125	38	9.0	55	2	I40315	filamentous hemagg	198	36.5	8.6	85	1	R86703	hypothetical prote
126	38	9.0	58	2	A45824	hypothetical prote	199	36.5	8.6	86	2	H87348	hypothetical prote
127	38	9.0	59	1	FEM2B	ferredoxin 2(4Fe-4	200	36.5	8.6	90	2	F82775	hypothetical prote
128	38	9.0	62	2	AB2131	hypothetical prote	201	36	8.5	47	2	B96949	hypothetical prote
129	38	9.0	63	1	FEDV2V	ferredoxin (4Fe-4S	202	36	8.5	50	2	T29687	hypothetical prote
130	38	9.0	64	2	D82849	hypothetical prote	203	36	8.5	52	2	H82829	hypothetical prote
131	38	9.0	64	2	D97731	hypothetical prote	204	36	8.5	55	2	JU0126	hypothetical prote
132	38	9.0	67	2	F90897	hypothetical prote	205	36	8.5	56	2	F95077	hypothetical prote
133	38	9.0	67	2	A85720	hypothetical prote	206	36	8.5	56	2	C72610	hypothetical prote
134	38	9.0	67	2	H64308	hypothetical prote	207	36	8.5	56	2	B82638	hypothetical prote
135	38	9.0	71	2	D98035	hypothetical prote	208	36	8.5	59	2	B97945	hypothetical prote
136	38	9.0	72	2	G64024	hypothetical prote	209	36	8.5	59	2	H69463	hypothetical prote
137	38	9.0	74	2	T26580	hypothetical prote	210	36	8.5	60	1	N1EPLJ	short neurotoxin 1
138	38	9.0	78	2	F95341	hypothetical prote	211	36	8.5	60	2	J50583	hypothetical 7K pr
139	38	9.0	78	2	F50085	hypothetical prote	212	36	8.5	60	2	A82568	hypothetical prote
140	38	9.0	82	2	S34332	protein [imported	213	36	8.5	61	2	A82932	hypothetical prote
141	38	9.0	84	2	S49416	major carboxysome	214	36	8.5	61	1	OSWU7C	cytochrome-c oxida
142	38	9.0	88	2	G87357	conserved hypothet	215	36	8.5	63	1	A18182	hypothetical prote
143	37.5	8.9	51	2	A42443	chitinase-related	216	36	8.5	63	2	A13360	hypothetical prote
144	37.5	8.9	52	2	S74783	hypothetical prote	217	36	8.5	64	2	B93398	protein [imported
145	37.5	8.9	66	2	A13361	hypothetical prote	218	36	8.5	65	2	T36228	hypothetical prote
146	37.5	8.9	66	2	A13361	hypothetical prote	219	36	8.5	67	2	A25590	1-aminocyclopropan
147	37.5	8.9	74	2	H70944	hypothetical prote	220	36	8.5	67	2	S20920	hypothetical prote
148	37.5	8.9	74	2	T48037	hypothetical prote	221	36	8.5	68	2	T30399	hypothetical prote
149	37.5	8.9	87	2	B64217	ribosomal protein	222	36	8.5	72	2	G87421	hypothetical prote
150	37.5	8.9	87	2	T45322	hypothetical prote	223	36	8.5	74	2	S34122	neurofilament prot
151	37.5	8.9	88	2	D82915	ribosomal protein	224	36	8.5	74	2	AH3104	conserved hypothet
152	37.5	8.9	89	2	G48909	dopamine receptor	225	36	8.5	75	2	AF0231	hypothetical prote
153	37	8.7	41	2	T03647	probable ribosomal	226	36	8.5	75	2	AC3575	hypothetical prote
154	37	8.7	53	1	LNLDAF	lectin alpha chain	227	36	8.5	78	2	C98182	hypothetical prote
155	37	8.7	53	2	A96216	hypothetical prote	228	36	8.5	80	2	S02067	glycerol kinase (E
156	37	8.7	54	2	C31436	ovomucoid, third d	229	36	8.5	80	2	AG9816	thiol-specific ant
157	37	8.7	54	2	G51587	ovomucoid, third d	230	36	8.5	82	2	A82307	ribosomal protein
158	37	8.7	58	2	A32071	T-cell receptor al	231	36	8.5	84	1	NTSR2N	Na+-channel-blocki
159	37	8.7	58	2	N1EPLV	short neurotoxin 1	232	36	8.5	84	2	T14543	peroxidase (EC 1.1
160	37	8.7	62	1	I36921	gene MHC DQ-alpha	233	36	8.5	84	2	H87242	hypothetical prote
161	37	8.7	62	2	I61802	gene MHC DQ-alpha	234	36	8.5	84	2	G90729	conserved hypothet
162	37	8.7	62	2	A82857	hypothetical prote	235	36	8.5	85	2	S30576	Ig light chain - r
163	37	8.7	64	2	AD2881	hypothetical prote	236	36	8.5	85	2	B35690	mucin 3 (clone SIB
164	37	8.7	66	2	S26378	cold shock protein	237	36	8.5	85	2	B37402	neurophysin - babo
165	37	8.7	66	2	H72738	hypothetical prote	238	36	8.5	86	2	T43977	hypothetical prote
166	37	8.7	66	2	AH0343	hypothetical prote	239	36	8.5	86	2	D69190	hypothetical prote
167	37	8.7	68	2	G69209	histone-related pr	240	36	8.5	89	2	AF0312	probable exported
168	37	8.7	73	2	A80258	hypothetical phase	241	35.5	8.4	49	2	T45355	hypothetical prote
169	37	8.7	74	2	T03188	hypothetical prote	242	35.5	8.4	55	2	D82595	hypothetical prote
170	37	8.7	78	2	H82796	hypothetical prote	243	35.5	8.4	61	2	C69260	hypothetical prote
171	37	8.7	78	2	G82593	hypothetical prote	244	35.5	8.4	61	2	S32030	Sp12 protein homol
172	37	8.7	79	2	G82593	conserved hypothet	245	35.5	8.4	62	2	S13389	lectin IV - Griffo
173	37	8.7	80	2	D72808	gp68.1 protein - M	246	35.5	8.4	62	2	AB1734	hypothetical prote
174	37	8.7	84	2	S76031	hypothetical prote	247	35.5	8.4	62	2	A11364	hypothetical prote
175	37	8.7	84	2	S76443	hypothetical prote	248	35.5	8.4	63	2	AH2565	hypothetical prote

249	35.5	8.4	65	2	A95330	hypothetical prote
250	35.5	8.4	69	2	A56378	GTP-binding regula
251	35.5	8.4	69	2	D97809	hypothetical prote
252	35.5	8.4	71	2	T09587	metallothionein-li
253	35.5	8.4	75	2	S07730	H+-transporting tw
254	35.5	8.4	75	2	T04547	hypothetical prote
255	35.5	8.4	78	2	E84686	hypothetical prote
256	35.5	8.4	79	2	F97432	hypothetical prote
257	35.5	8.4	82	2	F90468	copper binding pro
258	35.5	8.4	82	2	A0617	excisionase [ampr
259	35.5	8.4	83	2	B4679	hypothetical prote
260	35.5	8.4	84	2	D70611	hypothetical prote
261	35.5	8.4	86	2	H96679	hypothetical prote
262	35.5	8.4	87	2	S78075	hypothetical prote
263	35.5	8.4	87	2	D56273	probable transposa
264	35.5	8.4	88	2	C90178	ISU ribosomal prot
265	35.5	8.4	88	2	T29934	hypothetical prote
266	35.5	8.4	89	2	A43854	major merozoite su
267	35.5	8.4	89	2	T17956	hypothetical prote
268	35.5	8.4	89	2	H69871	hypothetical prote
269	35.5	8.4	89	2	D82528	hypothetical prote
270	35	8.3	40	2	A29184	hypothetical prote
271	35	8.3	50	2	F72765	vitellogenin - tur
272	35	8.3	53	2	S56726	probable ribosomal
273	35	8.3	54	2	S72975	calcium-dependent
274	35	8.3	55	2	A23382	glutamate decarbox
275	35	8.3	61	2	A81781	hypothetical prote
276	35	8.3	62	2	I47018	probable membrane
277	35	8.3	63	2	F59147	antigen WCL [simil
278	35	8.3	63	2	B5879	conotoxin p5.1 pre
279	35	8.3	68	2	G59093	cannabinoid recept
280	35	8.3	68	2	A28208	histone HmtA2 - Me
281	35	8.3	69	2	T18075	hypothetical prote
282	35	8.3	70	1	E18075	hypothetical prote
283	35	8.3	70	1	E18075	eglin C - medicina
284	35	8.3	72	2	A80023	hypothetical prote
285	35	8.3	72	2	T08871	interleukin 2 - mo
286	35	8.3	72	2	T03195	hypothetical prote
287	35	8.3	72	2	G71355	probable ribosomal
288	35	8.3	73	2	A29172	cathepsin B-like c
289	35	8.3	75	2	S17341	mating pheromone E
290	35	8.3	75	2	T24847	maternal pheromone E
291	35	8.3	75	2	T06954	hypothetical prote
292	35	8.3	77	2	AC3626	hypothetical prote
293	35	8.3	77	2	D97063	hypothetical prote
294	35	8.3	81	2	T18102	uncharacterized pr
295	35	8.3	82	2	F82644	hypothetical prote
296	35	8.3	82	2	F82644	hypothetical prote
297	35	8.3	83	2	JW0086	trichothecene 3-O-
298	35	8.3	83	2	T76476	hypothetical prote
299	35	8.3	84	2	T13498	NADH2 dehydrogenas
300	35	8.3	84	2	A41879	pvuII restriction
301	35	8.3	84	2	G97294	probable RNA-bind
302	35	8.3	85	2	T30462	hypothetical prote
303	35	8.3	85	2	A12765	hypothetical prote
304	35	8.3	85	2	AG3108	hypothetical prote
305	35	8.3	85	2	T13183	molybdopterin (mpt
306	35	8.3	88	2	A48435	repressor protein
307	35	8.3	89	2	A48435	cysteine proteinas
308	35	8.3	90	2	F90762	hypothetical prote
309	35	8.3	90	2	G85625	probable copy cont
310	34.5	8.2	43	2	B57484	hypothetical prote
311	34.5	8.2	43	2	A97416	cell division cont
312	34.5	8.2	45	2	T35464	hypothetical prote
313	34.5	8.2	50	2	S50105	ATP synthase alpha
314	34.5	8.2	55	2	A69152	DNA-dependent RNA
315	34.5	8.2	57	2	C41476	ribonucleoside-dip
316	34.5	8.2	58	2	S30499	protein-tyrosine k
317	34.5	8.2	58	2	D81150	hypothetical prote
318	34.5	8.2	60	2	D81150	hypothetical prote
319	34.5	8.2	61	2	A35567	photosystem I 14k
320	34.5	8.2	61	2	T21264	hypothetical prote
321	34.5	8.2	64	2	D46522	T-cell receptor et
322	34.5	8.2	64	2	H98033	hypothetical prote
323	34.5	8.2	66	2	E41991	S2 protein - equin
324	34.5	8.2	67	2	T36000	hypothetical prote
325	34.5	8.2	68	2	AD1306	probable mercuric
326	34.5	8.2	68	2	E82008	hypothetical prote
327	34.5	8.2	71	2	E98040	degenerative trans
328	34.5	8.2	72	2	S58489	hypothetical prote
329	34.5	8.2	72	2	I67891	yHDL protein - Esc
330	34.5	8.2	72	2	AE1875	hypothetical prote
331	34.5	8.2	73	2	D64683	ribosomal protein
332	34.5	8.2	75	2	JE0035	hypothetical 9K pr
333	34.5	8.2	76	2	B33335	integrin pS2 m8 al
334	34.5	8.2	77	2	B82441	hypothetical prote
335	34.5	8.2	77	2	T16222	hypothetical prote
336	34.5	8.2	78	2	AI0451	hypothetical prote
337	34.5	8.2	79	2	B95867	conserved hypotet
338	34.5	8.2	80	2	AF3364	hypothetical prote
339	34.5	8.2	81	2	T01671	vpu protein - huma
340	34.5	8.2	81	2	E59093	hypothetical prote
341	34.5	8.2	83	2	B95974	probable transcrip
342	34.5	8.2	84	2	E75545	conserved hypotet
343	34.5	8.2	85	2	T47903	ribosomal protein
344	34.5	8.2	85	2	B24886	P(-)lrl(-) fibrin
345	34.5	8.2	85	2	F55171	conserved hypotet
346	34.5	8.2	85	2	F98037	conserved hypotet
347	34.5	8.2	86	2	S50532	ribosomal protein
348	34.5	8.2	86	2	S55979	hypothetical prote
349	34.5	8.2	86	2	T03007	hypothetical prote
350	34.5	8.2	87	2	H91296	hypothetical prote
351	34.5	8.2	87	2	G71107	hypothetical prote
352	34.5	8.2	87	2	B86138	hypothetical prote
353	34.5	8.2	87	2	AD2239	hypothetical prote
354	34.5	8.2	89	2	A95026	ribosomal protein
355	34.5	8.2	89	2	B97897	30S ribosomal prot
356	34.5	8.2	89	2	D82717	30S ribosomal prot
357	34.5	8.2	89	2	T50144	intermembrane prot
358	34.5	8.2	90	2	A96505	probable glycine-r
359	34.5	8.2	90	2	D38179	hypothetical prote
360	34.5	8.2	90	2	T49083	hypothetical prote
361	34	8.0	42	2	I65746	trioponyosin - huma
362	34	8.0	44	2	A95166	hypothetical prote
363	34	8.0	44	2	E84151	hypothetical prote
364	34	8.0	46	2	G99537	gene 1.1 protein -
365	34	8.0	50	2	E90557	hypothetical prote
366	34	8.0	50	2	C72528	probable ribosomal
367	34	8.0	51	2	E2807	gp62 protein - Myc
368	34	8.0	61	2	F84126	4-oxalocrotonate t
369	34	8.0	62	2	S31217	cathepsin B-like c
370	34	8.0	62	2	AD3240	hypothetical prote
371	34	8.0	64	2	S32026	Sp12 protein homol
372	34	8.0	64	2	H95284	hypothetical prote
373	34	8.0	65	1	NTSR3C	neurotoxin 3 - bar
374	34	8.0	65	2	E82547	hypothetical prote
375	34	8.0	65	2	AF3177	conserved hypotet
376	34	8.0	66	1	NTSR2C	neurotoxin 2 - bar
377	34	8.0	66	2	G82491	hypothetical prote
378	34	8.0	66	2	T12847	hypothetical prote
379	34	8.0	68	2	JQ0099	hypothetical 7K pr
380	34	8.0	69	2	A61623	juvenile-hormone a
381	34	8.0	69	2	T03361	gene e6 protein -
382	34	8.0	69	2	AB1333	hypothetical prote
383	34	8.0	69	2	AB1704	hypothetical prote
384	34	8.0	71	2	T31215	hypothetical prote
385	34	8.0	72	2	AF2303	major carboxysome
386	34	8.0	72	2	S49415	hypothetical prote
387	34	8.0	72	2	S54448	hypothetical prote
388	34	8.0	72	2	B70517	lactoylglutathione
389	34	8.0	72	2	AH3473	aldehyde dehydroge
390	34	8.0	73	2	D81394	long neurotoxin 1
391	34	8.0	74	1	N2KF1U	hypothetical prote
392	34	8.0	74	2	F96670	hypothetical early
393	34	8.0	75	2	D61399	hypothetical prote
394	34	8.0	75	2	AG1917	hypothetical prote

395	34	8.0	76	2	I59567	histone H2A - rat	468	33.5	7.9	84	2	S58483	hypothetical prote	
396	34	8.0	77	2	C42519	A19L 8.3K protein	469	33.5	7.9	84	2	D87520	hypothetical prote	
397	34	8.0	77	2	E97137	hypothetical prote	470	33.5	7.9	84	2	F84388	hypothetical prote	
398	34	8.0	78	2	S03233	hypothetical prote	471	33.5	7.9	84	2	T51757	hypothetical prote	
399	34	8.0	78	2	S33333	hypothetical prote	472	33.5	7.9	84	2	C97194	hypothetical prote	
400	34	8.0	79	2	C03448	T-cell receptor al	473	33.5	7.9	84	2	AD1593	B. subtilis PBSX p	
401	34	8.0	80	2	S95994	hypothetical prote	474	33.5	7.9	85	2	T16739	hypothetical prote	
402	34	8.0	81	2	T17606	hypothetical prote	475	33.5	7.9	86	2	F87604	hypothetical prote	
403	34	8.0	81	2	T38051	probable glycoprot	476	33.5	7.9	88	2	T25449	probable flagellar	
404	34	8.0	82	2	G81477	BoIA/rIraA family p	477	33.5	7.9	88	2	AH0088	hypothetical prote	
405	34	8.0	82	2	D82020	hypothetical prote	478	33.5	7.9	89	1	W4B957	hypothetical prote	
406	34	8.0	83	2	S78469	Ig kappa chain v r	479	33.5	7.9	89	2	T03951	hypothetical prote	
407	34	8.0	84	1	W4WL	E4 protein - human	480	33	7.8	41	2	A59149	sigma-conotoxin GV	
408	34	8.0	84	2	S34099	Ig kappa chain v r	481	33	7.8	42	2	AD2520	hypothetical prote	
409	34	8.0	84	2	S74779	hypothetical prote	482	33	7.8	43	2	S35378	cysteine proteases	
410	34	8.0	86	1	D46335	W protein - Maedi/	483	33	7.8	47	2	H84193	30S ribosomal prot	
411	34	8.0	86	2	S16834	Ig kappa chain v r	484	33	7.8	53	2	B56557	hypothetical prote	
412	34	8.0	86	2	S16836	Ig kappa chain v r	485	33	7.8	54	2	S45255	PIEMP2/ME5A (clone	
413	34	8.0	86	2	S16827	Ig kappa chain v r	486	33	7.8	54	2	PT0183	YJX protein homol	
414	34	8.0	86	2	S16839	Ig kappa chain v r	487	33	7.8	55	2	170665	MHC HLA-A24 cell s	
415	34	8.0	86	2	S16828	Ig kappa chain v r	488	33	7.8	56	2	H82847	protein-tyrosine k	
416	34	8.0	86	2	S16835	Ig kappa chain v r	489	33	7.8	56	2	C84389	hypothetical prote	
417	34	8.0	86	2	S16824	Ig kappa chain v r	490	33	7.8	58	2	C97907	hypothetical prote	
418	34	8.0	86	2	S16829	Ig kappa chain v r	491	33	7.8	58	2	AG2271	hypothetical prote	
419	34	8.0	86	2	T00505	hypothetical prote	492	33	7.8	59	2	S18017	CAB/ELIP/HLIP supe	
420	34	8.0	87	2	S34098	Ig kappa chain v r	493	33	7.8	60	2	I59650	flocculant active	
421	34	8.0	87	2	A38725	transferrin - shee	494	33	7.8	61	2	S60796	class II histocomp	
422	34	8.0	87	2	H70079	hypothetical prote	495	33	7.8	61	2	D83531	M protein precurs	
423	34	8.0	87	2	T03193	hypothetical prote	496	33	7.8	62	2	T23216	carbon storage reg	
424	34	8.0	87	2	AF3030	hypothetical prote	497	33	7.8	62	2	H48411	hypothetical prote	
425	34	8.0	88	1	O6B977	gene 6.7 protein -	498	33	7.8	64	2	B55007	herculin homolog	
426	34	8.0	88	2	D90957	hypothetical prote	499	33	7.8	65	2	D84156	hypothetical prote	
427	34	8.0	88	2	A85492	unknown protein en	500	33	7.8	65	2	A94455	hypothetical prote	
428	34	8.0	88	2	A97936	conserved hypotet	501	33	7.8	66	2	B97981	hypothetical prote	
429	34	8.0	90	1	WNB936	gene m protein - p	502	33	7.8	66	2	I54475	hypothetical prote	
430	34	8.0	90	2	G71568	ribosomal protein	503	33	7.8	67	2	C59969	HLA-DNA-related sm	
431	34	8.0	90	2	E81260	nifu protein homol	504	33	7.8	68	2	I34256	hypothetical prote	
432	34	8.0	90	2	G84257	hypothetical prote	505	33	7.8	68	2	S54439	hemusrv operon pro	
433	34	8.0	90	2	F97577	hypothetical prote	506	33	7.8	70	2	G84810	probable proteinas	
434	34	8.0	90	2	A82798	hypothetical prote	507	33	7.8	71	2	T03353	gene c14 protein -	
435	33.5	7.9	45	2	D33982	ribosomal protein	508	33	7.8	72	2	A69292	archaeal histone A	
436	33.5	7.9	50	2	T11615	hypothetical prote	509	33	7.8	73	2	A27497	hypothetical prote	
437	33.5	7.9	52	2	S19574	hypothetical prote	511	33	7.8	74	2	I45714	Dpp receptor TKV,	
438	33.5	7.9	54	2	B90538	hypothetical prote	512	33	7.8	74	2	B75199	alpha-amylose inhi	
439	33.5	7.9	59	2	B93795	hypothetical prote	513	33	7.8	75	2	S10358	T-cell receptor be	
440	33.5	7.9	64	2	D97915	hypothetical prote	514	33	7.8	76	2	PC4256	protein-tyrosine k	
441	33.5	7.9	65	2	AB3381	heavy metal bindin	515	33	7.8	76	2	I52570	histone H3.1 - alf	
442	33.5	7.9	66	2	D98068	degenerate transpo	516	33	7.8	76	2	D82844	glycophorin - huma	
443	33.5	7.9	67	2	T16372	hypothetical prote	517	33	7.8	76	2	S75641	carbon storage reg	
444	33.5	7.9	67	2	F81618	hypothetical prote	518	33	7.8	77	2	S48011	hypothetical prote	
445	33.5	7.9	68	2	T11837	H+-transporting tw	519	33	7.8	77	2	G82169	phage shock protei	
446	33.5	7.9	68	2	S72596	hypothetical prote	520	33	7.8	77	2	S42162	hypothetical PHO80	
447	33.5	7.9	68	2	AE3561	hypothetical prote	521	33	7.8	78	1	A34931	glycophorin E prec	
448	33.5	7.9	69	2	B91001	hypothetical prote	522	33	7.8	78	1	F64439	DNA-directed RNA p	
449	33.5	7.9	74	2	H82816	omega-conotoxin GV	523	33	7.8	78	2	C75347	hypothetical prote	
450	33.5	7.9	75	2	A12832	kappa-casein - wat	524	33	7.8	80	1	C69782	conserved hypotet	
451	33.5	7.9	75	2	T06508	trbk protein - Ent	525	33	7.8	80	2	A89860	hypothetical prote	
452	33.5	7.9	75	2	I62139	gene Midl protein	526	33	7.8	80	2	D50965	hypothetical prote	
453	33.5	7.9	76	2	A11907	hypothetical prote	527	33	7.8	80	2	D85813	hypothetical prote	
454	33.5	7.9	77	2	D91137	hypothetical prote	528	33	7.8	80	2	F84959	hypothetical prote	
455	33.5	7.9	77	2	S07311	hypothetical prote	529	33	7.8	80	2	AF2420	hypothetical prote	
456	33.5	7.9	79	2	H75498	hypothetical prote	530	33	7.8	81	2	H97567	conserved hypotet	
457	33.5	7.9	79	2	T46591	hypothetical prote	531	33	7.8	81	2	S28280	hypothetical prote	
458	33.5	7.9	80	2	C96523	FLI17.12 (impor	532	33	7.8	82	2	H86302	probable 50S ribos	
459	33.5	7.9	80	2	S59544	stress-induced pro	533	33	7.8	83	2	C41378	hypothetical prote	
460	33.5	7.9	81	2	D95065	conserved domain p	534	33	7.8					
461	33.5	7.9	81	2	E97932	hypothetical prote	535	33	7.8					
462	33.5	7.9	81	2	E97932	hypothetical prote	536	33	7.8					
463	33.5	7.9	82	2	S78241	photosystem I iron	537	33	7.8					
464	33.5	7.9	82	2	F84057	hypothetical prote	538	33	7.8					
465	33.5	7.9	83	2	S62705	ribosomal protein	539	33	7.8					
466	33.5	7.9	83	2	B90338	hypothetical prote	540	33	7.8					
467	33.5	7.9	83	2	AC3327	hypothetical prote								

541	33	7.8	83	2	T26545	hypothetical prote	614	32.5	7.7	87	2	G65281	hypothetical prote
542	33	7.8	84	2	S78732	protein YMR122w-a	615	32.5	7.7	87	2	T00775	hypothetical prote
543	33	7.8	86	2	I56980	arginine decarboxy	616	32.5	7.7	87	2	T08641	hypothetical prote
544	33	7.8	86	2	C69752	hypothetical prote	617	32.5	7.7	88	2	D97092	hypothetical prote
545	33	7.8	87	1	E64990	Napd protein - Esc	618	32.5	7.7	89	2	E71977	ribosomal protein
546	33	7.8	87	2	S16842	Ig kappa chain v r	619	32.5	7.7	89	2	D64529	hypothetical prote
547	33	7.8	87	2	S40221	Glycoprotein B - h	620	32.5	7.7	89	2	D83372	hypothetical prote
548	33	7.8	87	2	B85860	Napd protein B - Esc	621	32.5	7.7	89	2	F70953	nonhistone chromos
549	33	7.8	87	2	S41863	gene NS-3 protein	622	32.5	7.7	90	1	NSCH77	phosphocarrrier pro
550	33	7.8	87	2	S73830	hypothetical prote	623	32.5	7.7	90	2	I76722	hypothetical prote
551	33	7.8	87	2	T48942	hypothetical prote	624	32.5	7.7	90	2	E91139	phosphocarrrier pro
552	33	7.8	88	2	H91015	Napd protein - Esc	625	32.5	7.7	90	2	F84289	hypothetical prote
553	33	7.8	88	2	AD0910	conserved hypothet	626	32.5	7.7	90	2	T46580	hypothetical prote
554	33	7.8	88	2	AP3180	agrobacterium viru	627	32.5	7.7	90	2	AE0991	hypothetical prote
555	33	7.8	88	2	AF2243	hypothetical prote	628	32.5	7.7	90	2	H85984	hypothetical prote
556	33	7.8	89	2	S07991	vpr protein - simi	629	32.5	7.7	90	2	D97689	hypothetical prote
557	33	7.8	89	2	B86741	hypothetical prote	630	32.5	7.7	90	2	A12914	hypothetical prote
558	33	7.8	90	2	C83960	ribosomal protein	631	32	7.6	91	2	H64039	probable membrane
559	33	7.8	90	2	E82456	conserved hypothet	632	32	7.6	91	2	A16039	phosphocarrrier pro
560	32.5	7.7	40	2	F87419	hypothetical prote	633	32	7.6	43	2	S35777	2'-aminobiphenyl-2
561	32.5	7.7	47	2	A56778	recombination prot	634	32	7.6	43	2	H82619	probable membrane
562	32.5	7.7	52	2	B81392	50S ribosomal prot	635	32	7.6	44	2	C30609	phosphocarrrier pro
563	32.5	7.7	52	2	I37971	zinc finger protei	636	32	7.6	45	2	A84042	hypothetical prote
564	32.5	7.7	53	2	T36215	spdB protein - Str	637	32	7.6	48	2	A69510	hypothetical prote
565	32.5	7.7	56	2	S31198	metallothionein -	638	32	7.6	48	2	E95852	hypothetical prote
566	32.5	7.7	57	2	I54529	major histocompat	639	32	7.6	49	2	I48681	185 kDa glycophos
567	32.5	7.7	57	2	C97840	hypothetical prote	640	32	7.6	52	2	S00580	1-5-phosphoribosy
568	32.5	7.7	61	2	AC3158	hypothetical prote	641	32	7.6	54	2	H31442	ovomucoid, third d
569	32.5	7.7	62	2	A45315	H <sup>+</sup> -transporting tw	642	32	7.6	54	2	I31444	ovomucoid, third d
570	32.5	7.7	62	2	S78537	tnpW protein - Clo	643	32	7.6	54	2	B36943	chemoreceptor prot
571	32.5	7.7	63	2	S29177	high-molecular-we	644	32	7.6	55	1	FEPE	chemoreceptor prot
572	32.5	7.7	64	2	T12941	hypothetical prote	645	32	7.6	55	2	AG2267	ferredoxin 2[4Fe-4
573	32.5	7.7	65	2	A95329	3C-like serine pro	646	32	7.6	56	2	G72355	hypothetical prote
574	32.5	7.7	65	2	S29904	hypothetical prote	647	32	7.6	56	2	C72613	hypothetical prote
575	32.5	7.7	66	2	A31395	hypothetical prote</							

687	32	7.6	68	2	H95264	hypothetical prote	760	32	7.6	86	2	S40226	glycoprotein B - h
688	32	7.6	68	2	Ad1678	probable mercuic	761	32	7.6	86	2	S76775	hypothetical prote
689	32	7.6	69	1	QOE67	yebB protein - Esc	762	32	7.6	86	2	T01496	hypothetical prote
690	32	7.6	69	2	G80713	hypothetical prote	763	32	7.6	86	2	E97027	probable transcrip
691	32	7.6	69	2	G85563	hypothetical prote	764	32	7.6	86	2	AF3172	DNA binding protei
692	32	7.6	69	2	H81902	hypothetical prote	765	32	7.6	87	2	S08401	neurotoxin homolog
693	32	7.6	69	2	J01703	hypothetical prote	766	32	7.6	87	2	JN0669	Na+-channel-blocki
694	32	7.6	69	2	C96021	hypothetical prote	767	32	7.6	87	2	JN0672	neurotoxin 4 precu
695	32	7.6	69	2	Ad1147	hypothetical prote	768	32	7.6	87	2	S40232	glycoprotein B - h
696	32	7.6	69	2	Ad1506	hypothetical prote	769	32	7.6	87	2	S40220	glycoprotein B - h
697	32	7.6	70	1	TIPO1A	chymotrypsin inhib	770	32	7.6	87	2	S40222	glycoprotein B - h
698	32	7.6	70	2	G08845	M protein precursor	771	32	7.6	87	2	S40234	small nuclear ribo
699	32	7.6	70	2	G42524	hypothetical prote	772	32	7.6	87	2	S90161	hypothetical prote
700	32	7.6	70	2	H82954	GTP-binding regula	773	32	7.6	87	2	A82064	M-related prote
701	32	7.6	70	2	A42155	hypothetical prote	774	32	7.6	87	2	C75458	hypothetical prote
702	32	7.6	71	2	H91105	hypothetical prote	775	32	7.6	87	2	T03315	gene l8 protein -
703	32	7.6	71	2	C85951	hypothetical prote	776	32	7.6	88	2	H70932	hypothetical prote
704	32	7.6	71	2	C95925	hypothetical prote	777	32	7.6	88	2	T34439	hypothetical prote
705	32	7.6	71	2	C65079	hypothetical 8.1 k	778	32	7.6	88	2	AH1367	B subtilis YtnI p
706	32	7.6	71	2	D81389	hypothetical prote	779	32	7.6	88	2	AE2108	hydroxymethylbilan
707	32	7.6	72	2	A86664	hypothetical prote	780	32	7.6	89	2	S23698	C8-2 protein - hep
708	32	7.6	72	2	A95351	Protein imported	781	32	7.6	89	2	S10002	hypothetical prote
709	32	7.6	73	2	I36840	C18L protein var	782	32	7.6	89	2	D84256	hypothetical prote
710	32	7.6	73	2	T28476	hypothetical prote	783	32	7.6	89	2	D84360	hypothetical prote
711	32	7.6	73	2	T03199	hypothetical prote	784	32	7.6	89	2	H36943	chemoreceptor prot
712	32	7.6	73	2	T07244	hypothetical prote	785	32	7.6	89	2	E69943	hypothetical prote
713	32	7.6	74	2	S61501	flagellar biosynth	786	32	7.6	89	2	G71378	hypothetical prote
714	32	7.6	75	2	B69391	DNA-directed RNA p	787	32	7.6	89	2	E65143	hypothetical 10.3k
715	32	7.6	75	2	S73297	hypothetical prote	788	32	7.6	90	2	A60526	complement C3 - ax
716	32	7.6	75	2	D97813	hypothetical prote	789	32	7.6	90	2	AH0464	probable membrane
717	32	7.6	76	2	S60827	M protein precursor	790	32	7.6	90	2	E97482	hypothetical 8.2k
718	32	7.6	76	2	A69391	DNA-directed RNA p	791	32	7.6	90	2	F97574	hypothetical prote
719	32	7.6	76	2	A72318	hypothetical prote	792	32	7.6	90	2	AF2795	hypothetical prote
720	32	7.6	76	2	D44007	apoptoxin IV - tra	793	32	7.6	90	2	T35208	hypothetical prote
721	32	7.6	76	2	AE2700	conserved hypotet	794	31.5	7.4	45	2	I54375	gene NF2 protein -
722	32	7.6	77	1	B32824	rhodotorucine A pr	795	31.5	7.4	48	2	JC6008	hypothetical 5.3k
723	32	7.6	77	2	T18049	hypothetical prote	796	31.5	7.4	48	2	AH2633	hypothetical prote
724	32	7.6	77	2	A71204	probable DNA-direc	797	31.5	7.4	49	2	B97874	degenerate transpo
725	32	7.6	77	2	AE2966	hypothetical prote	798	31.5	7.4	50	2	I52702	estrogen receptor
726	32	7.6	77	2	AE2097	hypothetical prote	799	31.5	7.4	51	2	A11731	hypothetical gene
727	32	7.6	78	2	B18883	hypothetical prote	800	31.5	7.4	51	2	E97394	hypothetical prote
728	32	7.6	78	2	T03641	glutathione S-tran	801	31.5	7.4	52	2	C91112	hypothetical prote
729	32	7.6	79	2	A49475	cerebroside sulfat	802	31.5	7.4	52	2	AC2566	hypothetical prote
730	32	7.6	79	2	J02128	metallothionein	803	31.5	7.4	52	2	F95269	hypothetical prote
731	32	7.6	79	2	B83400	hypothetical prote	804	31.5	7.4	52	2	G84912	hypothetical prote
732	32	7.6	79	2	AG0240	hypothetical prote	805	31.5	7.4	56	2	WMH51	ydbF protein - Esc
733	32	7.6	80	2	D71553	hypothetical prote	806	31.5	7.4	57	1	WMH24	3a protein - avian
734	32	7.6	80	2	J07435	hypothetical prote	807	31.5	7.4	57	2	C97937	transposase, uncha
735	32	7.6	80	2	D70487	hypothetical prote	808	31.5	7.4	58	2	B97333	ferredoxin [import
736	32	7.6	80	2	A48996	B cell surface ant	809	31.5	7.4	59	2	T06016	hypothetical prote
737	32	7.6	80	2	F98324	hypothetical prote	810	31.5	7.4	60	2	A28859	pepsin A (EC 3.4.2
738	32	7.6	80	2	AH2958	hypothetical prote	811	31.5	7.4	60	2	AC1904	hypothetical prote
739	32	7.6	80	2	AE2207	cell growth regula	812	31.5	7.4	62	2	C49570	plasma membrane ca
740	32	7.6	81	2	F97116	ribosomal protein	813	31.5	7.4	62	2	E97764	hypothetical prote
741	32	7.6	81	2	F75409	hypothetical prote	814	31.5	7.4	64	2	S17526	aconitate hydratase
742	32	7.6	82	2	D75321	ribosomal protein	815	31.5	7.4	64	2	AF3568	hypothetical prote
743	32	7.6	82	2	D81056	hypothetical prote	816	31.5	7.4	65	2	D86788	hypothetical prote
744	32	7.6	82	2	C81016	hypothetical prote	817	31.5	7.4	65	2	H81238	hypothetical prote
745	32	7.6	82	2	F89987	hypothetical prote	818	31.5	7.4	66	2	G72804	gp41.1 protein - M
746	32	7.6	83	2	E90859	hypothetical prote	819	31.5	7.4	66	2	T50506	hypothetical prote
747	32	7.6	83	2	H81188	hypothetical prote	820	31.5	7.4	67	2	AH0254	probable phage pro
748	32	7.6	83	2	S29325	hypothetical prote	821	31.5	7.4	67	2	S09322	fms protein - mous
749	32	7.6	83	2	B85760	hypothetical prote	822	31.5	7.4	68	2	S09322	adenylate cyclase
750	32	7.6	83	2	H85523	hypothetical prote	823	31.5	7.4	69	2	PN0453	hypothetical prote
751	32	7.6	83	2	AE0554	conserved hypotet	824	31.5	7.4	69	2	D95250	antibiotic mersac
752	32	7.6	83	2	D64875	yciN protein - Esc	825	31.5	7.4	69	2	F83706	hypothetical prote
753	32	7.6	83	2	S01442	hypothetical Alzhe	826	31.5	7.4	70	2	AE1974	transcription regu
754	32	7.6	84	1	A45631	ribosomal protein	827	31.5	7.4	70	2	H95067	X4JM protein - Rhi
755	32	7.6	84	2	T03163	hypothetical prote	828	31.5	7.4	70	2	T28640	T-cell receptor et
756	32	7.6	85	2	F75523	hypothetical prote	829	31.5	7.4	71	2	B46522	hypothetical prote
757	32	7.6	85	2	T40431	hypothetical prote	830	31.5	7.4	71	2	F37840	ORF1 protein - Orf
758	32	7.6	85	2	E95976	hypothetical prote	831	31.5	7.4	71	2	AG2853	hypothetical prote
759	32	7.6	86	2	C28195	Ig kappa chain V r	832	31.5	7.4	71	2	AG1893	hypothetical prote

833	31.5	7.4	72	2	T37089	probable IS elemen	906	31	7.3	55	2	A25988	mannose/glucose-sp
834	31.5	7.4	72	2	S13000	phosphopyruvate hy	907	31	7.3	55	2	S34330	sox protein - less
835	31.5	7.4	72	2	A75099	hypothetical prote	908	31	7.3	56	2	B72804	gp36 protein - Myc
836	31.5	7.4	72	2	AB3436	hypothetical prote	909	31	7.3	56	2	JW0095	p5 protein - human
837	31.5	7.4	72	2	AF2874	hypothetical prote	910	31	7.3	56	2	F82830	hypothetical prote
838	31.5	7.4	73	2	S78719	protein YER091c-a	911	31	7.3	56	2	E82787	cytochrome c - Try
839	31.5	7.4	74	2	G83393	hypothetical prote	912	31	7.3	57	2	A61479	hypothetical prote
840	31.5	7.4	74	2	A81648	hypothetical prote	913	31	7.3	57	2	D69120	hypothetical prote
841	31.5	7.4	74	2	AH1590	hypothetical prote	914	31	7.3	57	2	P69183	hypothetical prote
842	31.5	7.4	74	2	AH2642	hypothetical prote	915	31	7.3	58	2	C81787	hypothetical prote
843	31.5	7.4	75	2	S23602	hypothetical prote	916	31	7.3	59	2	A27606	ig heavy chain V-a
844	31.5	7.4	75	2	E82761	conserved hypotet	917	31	7.3	59	2	G91095	hypothetical prote
845	31.5	7.4	75	2	A10753	conserved hypotet	918	31	7.3	59	2	C85941	hypothetical prote
846	31.5	7.4	76	2	T19771	hypothetical prote	919	31	7.3	60	2	T03816	hypothetical prote
847	31.5	7.4	77	2	AH1279	hypothetical prote	920	31	7.3	60	2	A41212	tick anticoagulant
848	31.5	7.4	77	2	AH1642	hypothetical prote	921	31	7.3	61	2	D84210	hypothetical prote
849	31.5	7.4	78	2	T30726	hypothetical prote	922	31	7.3	61	2	C75321	preprotein translo
850	31.5	7.4	78	2	AH1969	hypothetical prote	923	31	7.3	61	2	AF1344	B. subtilis ywmg p
851	31.5	7.4	79	1	A70312	conserved hypotet	924	31	7.3	61	2	A11714	B. subtilis ywmg p
852	31.5	7.4	80	2	T10823	antifungal protein	925	31	7.3	61	2	AB3182	hypothetical prote
853	31.5	7.4	80	2	E64307	ferredoxin - Metha	926	31	7.3	62	1	CCB017	ubiquinol-cytochro
854	31.5	7.4	80	2	T38250	hypothetical prote	927	31	7.3	62	2	C70057	4-oxalocrotonate t
855	31.5	7.4	80	2	C24956	hypothetical prote	928	31	7.3	62	2	S38578	probable serine/th
856	31.5	7.4	81	1	CD9880	conserved hypotet	929	31	7.3	62	2	C83944	hypothetical prote
857	31.5	7.4	81	2	S39511	photosystem I iron	930	31	7.3	62	2	C82647	hypothetical prote
858	31.5	7.4	81	2	T44031	hypothetical prote	931	31	7.3	63	1	Q0ECLM	yaia protein - Esc
859	31.5	7.4	82	2	T07076	metallothionein ty	932	31	7.3	63	2	AD0550	conserved hypotet
860	31.5	7.4	82	2	T07114	metallothionein-11	933	31	7.3	63	2	C85534	hypothetical prote
861	31.5	7.4	82	2	E83191	conserved hypotet	934	31	7.3	63	2	G90883	hypothetical prote
862	31.5	7.4	82	2	F71890	hypothetical prote	935	31	7.3	63	2	JC4002	carbomycin resista
863	31.5	7.4	83	2	AB2519	hypothetical prote	936	31	7.3	63	2	D86673	hypothetical prote
864	31.5	7.4	84	2	S03239	hypothetical prote	937	31	7.3	63	2	A59231	hypothetical prote
865	31.5	7.4	84	2	AG2301	hypothetical prote	938	31	7.3	63	2	S58623	hypothetical prote
866	31.5	7.4	85	2	AC2533	hypothetical prote	939	31	7.3	63	2	T33630	hypothetical prote
867	31.5	7.4	86	2	S72340	hypothetical prote	940	31	7.3	63	2	S29654	juvenile-hormone e
868	31.5	7.4	86	2	F64527	hypothetical prote	941	31	7.3	64	2	D84219	hypothetical prote
869	31.5	7.4	86	2	F71978	hypothetical prote	942	31	7.3	64	2	F97207	probable transcrip
870	31.5	7.4	87	2	A84679	hypothetical prote	943	31	7.3	65	2	JC2380	subtilisin inhibit
871	31.5	7.4	87	2	T03716	reverse transcript	944	31	7.3	66	2	H71940	copper-associated
872	31.5	7.4	87	2	C87533	hypothetical prote	945	31	7.3	66	2	S47027	DNA-directed RNA p
873	31.5	7.4	88	2	B97057	ribosomal protein	946	31	7.3	67	1	PWMS8	H+-transporting tw
874	31.5	7.4	88	2	A97425	hypothetical prote	947	31	7.3	67	2	C86830	hypothetical prote
875	31.5	7.4	88	2	AB1901	hypothetical prote	948	31	7.3	67	2	C81174	hypothetical prote
876	31.5	7.4	89	2	S13517	retinoic acid rece	949	31	7.3	67	2	S76880	small hypotet
877	31.5	7.4	89	2	G59730	holin component xh	950	31	7.3	67	2	T35269	milk band E protei
878	31.5	7.4	90	2	S03700	nonhistone chroms	951	31	7.3	68	2	A61511	17K common-antigen
879	31.5	7.4	90	2	AB9103	hypothetical prote	952	31	7.3	68	2	AE3441	proteinase inhibit
880	31	7.3	40	2	B81705	segment 12 protein	953	31	7.3	69	2	S66650	ribosomal protein
881	31	7.3	43	2	H27712	hypothetical prote	954	31	7.3	69	2	JQ1257	hypothetical 7.6K
882	31	7.3	44	2	S05348	nitrogenase (EC 1.	955	31	7.3	69	2	JC4731	mob protein 69 - E
883	31	7.3	44	2	A82865	hypothetical prote	956	31	7.3	69	2	G98318	hypothetical prote
884	31	7.3	45	2	C34689	calmodulin-related	957	31	7.3	69	2	AD2483	hypothetical prote
885	31	7.3	46	2	B29184	vitellogenin - duc	958	31	7.3	70	2	A30518	ig heavy chain V-A
886	31	7.3	46	2	E90676	hypothetical prote	959	31	7.3	70	2	B22286	ferredoxin - Therm
887	31	7.3	46	2	H85526	hypothetical prote	960	31	7.3	70	2	AB1103	mercury transport
888	31	7.3	47	2	S62326	cystatin C, paroti	961	31	7.3	70	2	E82148	conserved hypotet
889	31	7.3	47	2	E84357	50S ribosomal prot	962	31	7.3	70	2	T12849	hypothetical prote
890	31	7.3	48	2	B69173	ribosomal protein	963	31	7.3	70	2	AI2632	conserved hypotet
891	31	7.3	49	2	B82007	hypothetical prote	964	31	7.3	70	2	AG2846	cold shock protein
892	31	7.3	49	2	A82422	hypothetical prote	965	31	7.3	71	2	I49244	MI protein - mouse
893	31	7.3	50	2	J01600	repC' protein - Cl	966	31	7.3	71	2	F84312	hypothetical prote
894	31	7.3	50	2	A60745	major outer membra	967	31	7.3	71	2	JQ0147	hypothetical 7.7K
895	31	7.3	50	2	H96530	hypothetical prote	968	31	7.3	71	2	AC0229	probable membrane
896	31	7.3	51	2	S11815	opacity protein (c	969	31	7.3	71	2	H97623	hypothetical prote
897	31	7.3	51	2	AF2492	hypothetical prote	970	31	7.3	72	1	TIMB	trypsin inhibitor
898	31	7.3	52	2	T43384	hypothetical prote	971	31	7.3	72	2	S49186	ferredoxin 2[4Fe-4
899	31	7.3	53	2	S44404	NADH2 dehydrogenas	972	31	7.3	72	2	S49021	retinoic acid rece
900	31	7.3	53	2	F82856	hypothetical prote	973	31	7.3	72	2	F10937	cellulase (EC 3.2.
901	31	7.3	53	2	AC3113	hypothetical prote	974	31	7.3	72	2	F95136	NADH-ferredoxin impo
902	31	7.3	54	2	I46202	galanin - dog (fra	975	31	7.3	72	2	D70167	hypothetical prote
903	31	7.3	54	2	S24179	histone H1.a, hepa	976	31	7.3	73	2	JC1066	trypsin inhibitor
904	31	7.3	54	2	H82718	hypothetical prote	977	31	7.3	73	2	JC1066	hypothetical prote
905	31	7.3	55	2	F70186	protein-tyrosine k	978	31	7.3	73	2	A84271	hypothetical prote

979 31 7.3 73 2 E84286  
 980 31 7.3 73 2 C64021  
 981 31 7.3 73 2 S12242 rap  
 982 31 7.3 73 2 E14L protein - var  
 983 31 7.3 73 2 E84048  
 984 31 7.3 74 2 S13515  
 985 31 7.3 74 2 S127539  
 986 31 7.3 74 2 F71080  
 987 31 7.3 74 2 C96982  
 988 31 7.3 74 2 E97870  
 989 31 7.3 75 2 A38646  
 990 31 7.3 75 2 T08193  
 991 31 7.3 75 2 T01621  
 992 31 7.3 76 2 C82122  
 993 31 7.3 76 2 T40983  
 994 31 7.3 76 2 C44007  
 995 31 7.3 76 2 A95337  
 996 31 7.3 77 2 D69432  
 997 31 7.3 77 2 T02773  
 998 31 7.3 78 2 E64104  
 999 31 7.3 78 2 A24352  
 1000 31 7.3 78 2 A75084

ALIGNMENTS

hypothetical protein ZC477.4 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
 C:Accession: T27603

Submitted to the EMBL Data Library, November 1995  
 A:Description: The sequence of C. elegans cosmid ZC477.  
 A:Reference number: Z20392  
 A:Accession: T27603  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-80 <DUZ>  
 A:Cross-references: EMBL:U040802; PIDN:AAA81504.1; CESP:ZC477.4  
 C:Genetics:  
 A:Gene: CESP:ZC477.4

Query Match 12.8%; Score 54; DB 2; Length 80;  
 Best Local Similarity 28.4%; Pred. No. 39;  
 Matches 23; Conservative 9; Mismatches 35; Indels 14; Gaps 2;

QY 6 CSSOSISPMRSISNSLVAMDFSGKSRVIENTEALSVAVEGLAWKKGCLRLGTHGS 65  
 DB 4 CSPLKILPGASSSSSSSTA-----SSQIRPFLSLSASLSEELRVEECGSPRVGAKES 56

QY 66 -----PTASSOSSATNNAI 79  
 DB 57 SFYCTEQPAQSSYSREDKLC 77

RESULT 2  
 DNA-directed RNA polymerase (EC 2.7.7.6) subunit N - Methanococcus jannaschii  
 C:Species: Methanococcus jannaschii  
 C:Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 23-Apr-1999  
 C:Accession: E64324  
 R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake,  
 rson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.  
 Science 273, 1058-1073, 1996  
 A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese,  
 A:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii  
 A:Reference number: A64300; MUID:96337999; PMID:8688087  
 A:Accession: E64324  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA  
 A:Residues: 1-76 <BUL>  
 A:Cross-references: GB:U67475; GB:L77117; NID:g1590930; PID:g1590941; TIGR:MJ0196; PI  
 C:Genetics:  
 A:Map position: FOR190573-190803  
 A:Start codon: GTG  
 C:Superfamily: DNA-directed RNA polymerase II chain RPB10  
 C:Keywords: nucleotidyltransferase; transcription

Query Match 12.2%; Score 51.5; DB 2; Length 76;  
 Best Local Similarity 30.0%; Pred. No. 71;  
 Matches 15; Conservative 10; Mismatches 18; Indels 7; Gaps 2;

QY 13 PMBSISNSLVAMDFSGKSRVI--ENPTAALSVAVEGLAWKKGCLRL 60  
 DB 7 PIRCFSGNVIAVEVEEYKELIKGENPKDVL-----DGLGKKYCCRRM 51

RESULT 3  
 AG3246  
 conserved hypothetical protein Atu6153 [imported] - Agrobacterium tumefaciens (strain  
 C:Species: Agrobacterium tumefaciens  
 C:Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 11-Jan-2002  
 C:Accession: AG3246  
 R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo  
 erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McCl  
 ; Karp, P.; Romero, P.; Zhang, S.  
 Science 294, 2317-2323, 2001  
 A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kam  
 ster, E.W.  
 A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.  
 A:Reference number: AB2577; PMID:11743193  
 A:Accession: AG3246  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-89 <KUR>  
 A:Cross-references: GB:AE008690; PIDN:AAL46389.1; PID:g17744182; GSPDB:GN00189  
 A:Experimental source: strain C58 (Dupont)  
 C:Genetics:  
 A:Gene: Atu6153  
 A:Genome: Plasmid

Query Match 11.3%; Score 48; DB 2; Length 89;  
 Best Local Similarity 41.2%; Pred. No. 2,2e+02;  
 Matches 14; Conservative 5; Mismatches 9; Indels 6; Gaps 1;

QY 28 SGQKSRVIENTEALSVAVEGLAWKKGCLRLG 61  
 DB 29 SQKRSRV-----PCSLAVISSMLHROKGAVRVG 56

RESULT 4  
 F95310  
 hypothetical protein Sma0736 [imported] - Sinorhizobium meliloti (strain 1021) magap1  
 C:Species: Sinorhizobium meliloti  
 C:Date: 24-Aug-2001 #sequence\_revision 24-Aug-2001 #text\_change 30-Sep-2001  
 C:Accession: F95310  
 R:Barrett, M.J.; Fisher, R.F.; Jones, T.; Komp, C.; Abola, A.P.; Barloy-Hubler, F.; B  
 ; Kalman, S.; Keating, D.H.; Palm, C.; Peck, M.C.; Surzycki, R.; Wells, D.H.; Yeh, K  
 proc. Natl. Acad. Sci. U.S.A. 98, 9883-9888, 2001  
 A:Title: Nucleotide sequence and predicted functions of the entire Sinorhizobium meli  
 A:Reference number: A95262; MUID:21396509; PMID:11481432  
 A:Accession: F95310  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-90 <KUR>  
 A:Cross-references: GB:AE006469; PIDN:AAK65048.1; PID:g14523480; GSPDB:GN00165  
 A:Experimental source: strain 1021, megaplasmid pSymA  
 R:Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubl  
 pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.  
 L.; Hyman, R.W.; Jones, T.  
 Science 293, 668-672, 2001  
 A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelau

hebaault, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.  
 A:Title: The composite genome of the legume symbiont *Sinorhizobium meliloti*.  
 A:Reference number: A96039; MUID:21368234; PMID:11474104  
 C:Contents: annotation  
 C:Genetics:  
 A:Gene: Sma0736  
 A:Genome: plasmid

Query Match 11.2%; Score 47.5; DB 2; Length 90;  
 Best Local Similarity 22.1%; Pred. No. 2.5e+02;  
 Matches 15; Conservative 15; Mismatches 19; Indels 19; Gaps 3;  
 QY 11 ISPMRISNSLVAMPDSQKSRVIENTEALSVAVEEGLW-----RRKGCLRL-GT 62  
 DB 11 VAPPKDLDDSTMSAADFA-----QLFGVYTOGGMSETFYERQADGIFRLSGT 59  
 QY 63 HGSPTASS 70  
 DB 60 RKTNTPS 67

RESULT 5  
 E64510  
 hypothetical protein MJEBL05 - Methanococcus jannaschii plasmid purB800  
 C:Species: Methanococcus jannaschii  
 C:Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 22-Oct-1999  
 C:Accession: E64510  
 R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake,  
 ; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.;  
 rson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.  
 Science 273, 1058-1073, 1996  
 A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C.  
 A:Title: Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*  
 A:Reference number: A64300; MUID:96337999; PMID:8688087  
 A:Accession: E64510  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-62 <BUL>  
 A:Cross-references: GB:L77118; MID:g1500644; TIGR:MJEBL05; PIDN:AAC37071.1; PID:g1500645  
 C:Genetics:  
 A:Map position: ECLFOR3265-3453  
 A:Genome: plasmid  
 A:Start codon: GTG  
 A:Note: This stable 58-kilobase pair plasmid is also designated ECL (large extrachromosomal

Query Match 11.0%; Score 46.5; DB 2; Length 62;  
 Best Local Similarity 28.6%; Pred. No. 2.1e+02;  
 Matches 12; Conservative 8; Mismatches 21; Indels 1; Gaps 1;  
 QY 15 RSISENSLVAMPDS-COKSRVIENTEALSVAVEEGLAWKK 55  
 DB 18 KVAERFLKDLSSQGDWKEIRAEAKKKLEEGIAWK 59

RESULT 6  
 S24471  
 gag polyprotein - human immunodeficiency virus type 1  
 C:Species: human immunodeficiency virus type 1, HIV-1  
 C:Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 26-Aug-1999  
 C:Accession: S24471; S24483  
 R:Salminen, M.  
 submitted to the EMBL Data Library, October 1991  
 A:Reference number: S24471  
 A:Accession: S24471  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-77 <SAL>  
 A:Cross-references: EMBL:211145; MID:g60073; PIDN:CAA77496.1; PID:g60074  
 C:Superfamily: AIDS-related virus gag polyprotein  
 C:Keywords: polyprotein

Query Match 10.9%; Score 46; DB 2; Length 77;  
 Best Local Similarity 23.4%; Pred. No. 3.1e+02;

Matches 15; Conservative 7; Mismatches 14; Indels 28; Gaps 2;  
 QY 29 GOKSRV-----ENPTEALSVAVEEG-----LAWRKGCRL 60  
 DB 1 GHKARYLAQAMSKATNAATIMMORGNFRNKTVCFCNCGKQGHIAARNCRARKKGCWKC 60  
 QY 61 GTHG 64  
 DB 61 GREG 64

RESULT 7  
 D82861  
 hypothetical protein XFa0004 [imported] - *Xylella fastidiosa* (strain 9a5c)  
 C:Species: *Xylella fastidiosa*  
 C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 20-Aug-2000  
 C:Accession: D82861  
 R:anonymous, The *Xylella fastidiosa* Consortium of the Organization for Nucleotide S  
 Nature 406, 151-157, 2000  
 A:Title: The genome sequence of the plant pathogen *Xylella fastidiosa*.  
 A:Reference number: A82515; MUID:20365717; PMID:10910347  
 A:Note: for a complete list of authors see reference number A59328 below  
 A:Accession: D82861  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-87 <SIN>  
 A:Cross-references: GB:AE003851; MID:g9112238; PIDN:AAF85573.1; GSPDB:GN00130; XFSC  
 R:Experimental source: strain 9a5c  
 R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga,  
 Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carr  
 as-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.  
 submitted to GenBank, June 2000  
 A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.;  
 J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.B.;  
 Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Marti  
 A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki,  
 F.G.; Nunes, L.R.; Oliveira, A.J.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri  
 Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sa  
 A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Si  
 M.Tsunako, M.H.; Valtada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.  
 A:Reference number: A59328  
 A:Contents: annotation  
 C:Genetics:  
 A:Gene: Xfa0004  
 A:Genome: plasmid  
 A:Note: plasmid pXF5.1

Query Match 10.9%; Score 46; DB 2; Length 87;  
 Best Local Similarity 28.1%; Pred. No. 3.6e+02;  
 Matches 16; Conservative 10; Mismatches 23; Indels 8; Gaps 3;  
 QY 14 MRSISENSLVAMPDF---SGOKSRVIENTEALSVAVEEGLAWRKGCRLGTHGSP 66  
 DB 1 MNTISESELA--DFVKNQAVKKLHIIQNDAEKYEIFAT--LTWKEGDWHLVTTTRCKP 53

RESULT 8  
 A82657  
 hypothetical protein XFL634 [imported] - *Xylella fastidiosa* (strain 9a5c)  
 C:Species: *Xylella fastidiosa*  
 C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 20-Aug-2000  
 C:Accession: A82657  
 R:anonymous, The *Xylella fastidiosa* Consortium of the Organization for Nucleotide S  
 Nature 406, 151-157, 2000  
 A:Title: The genome sequence of the plant pathogen *Xylella fastidiosa*.  
 A:Reference number: A82515; MUID:20365717; PMID:10910347  
 A:Note: for a complete list of authors see reference number A59328 below  
 A:Accession: A82657  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-65 <SIN>  
 A:Cross-references: GB:AE003990; GB:AE003849; MID:g9106683; PIDN:AAF84443.1; GSPDB:1  
 A:Experimental source: strain 9a5c



```

Query Match          10.8%;   Score 45.5;   DB 2;   Length 81;
Best Local Similarity    25.5%;   Pred. No. 3.8e+02;
Matches 14; Conservative      8; Mismatches 22; Indels 11; Gaps 1;

QY 9 QISPMRSEISSLVAMDFSGKSRVTIENTEALSVAVEGLAWRKKGCLRLGTH 63
DB 5 QKIPGVCKATEKLIMLGTYTIKSLKDANPAQMY-----EKELMRQH 48

RESULT 11
A:2960
C:ferredoxin 2[4Fe-4S] - Methanosarcina thermophila
C:Species: Methanosarcina thermophila
C>Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 13-Nov-1998
C:Accession: A42960
R:K.Clements, A.P.; Ferry, J.G.
J. Bacteriol. 174, 5244-5250, 1992
A:Title: Cloning, nucleotide sequence, and transcriptional analyses of the gene encoded
A:Reference number: A42960; MUID:92355496; PMID:1379583
A:Contents: TM-1
A:Accession:: A42960
A:Molecule type: DNA
A:Residues: 1-60 <CLE>
A:Note: sequence extracted from NCBI backbone (NCBIN:l10322, NCBIP:l10324)
C:Genetics:
A:Gene: fdxA
C:Superfamily: ferredoxin 2[4Fe-4S]; ferredoxin 2[4Fe-4S] homology
C:Keywords: 4Fe-4S; electron transfer; iron-sulfur protein; metalloprotein
F:3-59/Domain: ferredoxin 2[4Fe-4S] homology <FER>
F:10.13.16.51/Binding site: 4Fe-4S cluster (Cys) (covalent) #status predicted
F:20.41.44.47/Binding site: 4Fe-4S cluster (Cys) (covalent) #status predicted

Query Match          10.6%;   Score 45;   DB 2;   Length 60;
Best Local Similarity    42.9%;   Pred. No. 3e+02;
Matches 12; Conservative      7; Mismatches 9; Indels 0; Gaps 0;

QY 24 AMDFSQKRSRVIENTPEALSVAVEGLA 51
DB 7 ADECSGGSCVDCEPSEAITLDEKGI 34

RESULT 12
I53107
CD24 precursor - rat
C:Species: Rattus norvegicus (Norway rat)
C>Date: 02-Aug-1996 #sequence_revision 02-Aug-1996 #text_change 05-Nov-1999
C:Accession: I53107; S25146
R:Shirasawa, T.; Akashi, T.; Sakamoto, K.; Takahashi, H.; Maruyama, N.; Hirokawa, K.
Dev. Dyn. 198, 1-13, 1993
A:Title: Gene expression of CD24 core peptide molecule in developing brain and developm
A:Reference number: I53107; MUID:94122434; PMID:8292828
A:Accession: I53107
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-76 <RES>
A:Cross-references: EMBL:211663; MID:g55901; PIDN:CAA7731.1; PID:g55902
C:Keywords: phosphatidylinositol linkage

Query Match          10.6%;   Score 45;   DB 2;   Length 76;
Best Local Similarity    24.4%;   Pred. No. 4e+02;
Matches 19; Conservative     10; Mismatches 15; Indels 34; Gaps 4;

QY 6 CSSOSIPMSRISNSLVAMDFSGOKS-RVIENPTLEALSVAVEGLAWRKKGCLRLGTHG 64
DB 2 CNOTSVAP-----FSGNQSIAAPNPNTAVT-----RSGC-----55

QY 65 SPTASSOSSATNNMATHRS 82
DB 56 ---SSLSQTAGLLALSLS 70

RESULT 13

```

Tl7014  
 metallothionein-like protein AMT1 - apple tree  
 acc:Species: Malus domestica (apple tree)  
 date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 11-Jan-2000  
 accession: Tl7014  
 reid: S.J.; Ross, G.S.  
 physiol. plantarum 100, 183-189, 1997  
 title: Up-regulation of two CDNA clones encoding metallothionein-like proteins in apple  
 reference number: 218652  
 accession: Tl7014  
 status: preliminary; translated from GB/EMBL/DBJ  
 molecule type: mRNA

Best Local Similarity 25.4%; Pred No. 4.6e+02;  
Matches 18; Conservative 7; Mismatches 20; Indels 26; Gaps 4;

QY 53 RKGGLRLGTH 63  
DB 74 --RSC--VGTH 80

## RESULT 18

AB2271  
periplasmic mercuric ion binding protein [imported] - Nostoc sp. (strain PCC 7120)  
C:Species: Nostoc sp.  
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 30-Jun-2002  
C:Accession: AB2271  
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S. DNA Res. 8, 205-213, 2001  
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena sp. strain PCC 7120  
A:Reference number: AB1807; MUID:21595285; PMID:11759840  
A:Accession: AB2271  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-64 <KUR>  
A:Cross-references: GB:BA000019; PIDN:BA075420.1; PID:g17132855; GSPDB:GN00179  
A:Experimental source: strain PCC 7120  
C:Genetics:  
A:Gene: asl3721

Query Match 10.5%; Score 44.5; DB 2; Length 64;  
Best Local Similarity 16.7%; Pred. No. 3.7e+02;  
Matches 8; Conservative 20; Mismatches 17; Indels 3; Gaps 1;

QY 4 SGCSQSISPMRSISNSILVAMDFGSKSRVNIENPTALSVAVEGLA 51  
DB 13 SACANNITNAKTVDDAIVQAD---PQTLVNVVETQASSETSIIKDALA 57

## RESULT 19

S75293  
hypothetical protein srr2333 - Synecocystis sp. (strain PCC 6803)  
C:Species: Synecocystis sp.  
A:Variety: PCC 6803  
C:Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 08-Oct-1999  
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda, M. DNA Res. 3, 109-136, 1996  
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synecocystis sp. strain PCC 6803  
A:Reference number: S75293  
A:Accession: S75293  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-79 <KAN>  
A:Cross-references: EMBL:D90904; GB:AB001339; NID:g1652225; PIDN:BAAL7207.1; PID:d101794  
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

Query Match 10.5%; Score 44.5; DB 2; Length 79;  
Best Local Similarity 22.1%; Pred. No. 4.8e+02;  
Matches 15; Conservative 14; Mismatches 34; Indels 5; Gaps 1;

QY 16 SISNSLVAMDFGSKSRVNIENPTALSVAVEGLAWRKGCRLRGTGSPASSQSAT 75  
DB 7 SVGOLAVKILLGNHGOGLVNRLEAMGIIPDKPIQLLRKAGL----GGPLHLRIGSTT 61

QY 76 NNAIHRSQ 83  
DB 62 EVAMRSE 69

## RESULT 20

DB5807  
hypothetical protein 22988 [imported] - Escherichia coli (strain O157:H7, substrain EDLs)  
C:Species: Escherichia coli

C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 14-Sep-2001  
C:Accession: D85807  
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May, J.; Miller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamouitis, K.; Apoda, N. Nature 409, 529-533, 2001  
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
A:Reference number: A85480; MUID:21074935; PMID:11206551  
A:Accession: D85807  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-51 <STO>  
A:Cross-references: GB:AE005174; NID:g12516000; PIDN:AA056920.1; GSPDB:GN00145; UWGP:  
A:Experimental source: strain O157:H7, substrain EDL933  
C:Genetics:  
A:Gene: 22988

Query Match 10.4%; Score 44; DB 2; Length 51;  
Best Local Similarity 32.0%; Pred. No. 3.2e+02;  
Matches 8; Conservative 6; Mismatches 11; Indels 0; Gaps 0;

QY 37 NPTALSVAVEEGLAWRKKGCLRLG 61  
DB 20 SPAEIFMWTGPEVVSWRERAAALRS 44

## RESULT 21

B90959  
probable phage tail protein [imported] - Escherichia coli (strain O157:H7, substrain C:Species: Escherichia coli  
C:Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 18-Jul-2001  
C:Accession: B90959  
R:Havashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C. gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H. DNA Res. 8, 11-22, 2001  
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and g  
A:Reference number: A99629; MUID:21156231; PMID:11258796  
A:Accession: B90959  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-78 <HAY>  
A:Cross-references: GB:BA000007; PIDN:BA036065.1; PID:g13362110; GSPDB:GN00154  
A:Experimental source: strain O157:H7, substrain RMD 0509952  
C:Genetics:  
A:Gene: EC2642

Query Match 10.4%; Score 44; DB 2; Length 78;  
Best Local Similarity 32.0%; Pred. No. 5.3e+02;  
Matches 8; Conservative 6; Mismatches 11; Indels 0; Gaps 0;

QY 37 NPTALSVAVEEGLAWRKKGCLRLG 61  
DB 47 SPAEIFMWTGPEVVSWRERAAALRS 71

## RESULT 22

A75411  
hypothetical protein - Deinococcus radiodurans (strain R1)  
C:Species: Deinococcus radiodurans  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 17-Mar-2000  
C:Accession: A75411  
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J. S.; Smith, H.O.; Venter, J.C.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Science 286, 1571-1577, 1999  
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.  
A:Reference number: A75250; MUID:20036896; PMID:10567265  
A:Accession: A75411  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-78 <WHI>  
A:Cross-references: GB:AE001978; NID:g6459059; PIDN:AAF10892.1; PID:9645  
A:Experimental source: strain R1  
C:Genetics:



GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run On: March 28, 2003, 09:08:20 ; Search time 27.1181 Seconds  
(without alignments)  
181,900 Million cell updates/sec

Sequence: US-09-936-697-6  
Reference score: 493.0  
Sequence: 1 OGRSGSSQSSISPMRSISEN.....SPTASSQSSATNMAHRSQP 84

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 237916 seqs, 58723674 residues

Total number of hits satisfying chosen parameters: 29382

Minimum DB seq length: 40  
Maximum DB seq length: 90

Post-processing: Minimum Match 08

Maximum Match 100%  
Listing first 1000 summaries

Published Applications: AA  
1: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB pep.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB pep.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB pep.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB pep.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB pep.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB pep.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	54.5	12.9	60	9	US-09-925-299-1024
2	54.5	12.9	60	10	US-09-925-299-1024
3	51.5	12.2	84	10	US-09-908-711-103
4	51	12.1	84	10	US-09-864-761-47521
5	50.5	11.9	80	9	US-10-125-258-6
6	48.5	11.5	82	10	US-09-864-761-42490
7	47	11.1	89	9	US-09-764-872-393
8	46.5	11.0	78	10	US-09-764-887-218
9	46	10.9	68	9	US-10-079-854-131
10	46	10.9	68	10	US-09-764-878-131
11	46	10.9	68	10	US-09-764-860-468
12	46	10.9	69	9	US-10-092-154-700
13	46	10.8	69	10	US-09-764-847-700
14	45.5	10.8	56	10	US-09-925-299-1460
15	45.5	10.8	56	9	US-09-925-299-1460
16	45.5	10.8	58	9	US-09-925-299-1372
17	45.5	10.8	58	9	US-09-925-299-1437
18	45.5	10.8	58	9	US-09-925-299-1493
19	45.5	10.8	58	9	US-09-925-299-1528

20	45.5	10.8	58	10	US-09-925-301-1647	Sequence 1647, Ap
21	45.5	10.8	58	10	US-09-925-299-1372	Sequence 1372, Ap
22	45.5	10.8	58	10	US-09-925-299-1437	Sequence 1437, Ap
23	45.5	10.8	58	10	US-09-925-299-1493	Sequence 1493, Ap
24	45.5	10.8	58	10	US-09-925-299-1528	Sequence 1528, Ap
25	45.5	10.8	60	10	US-09-764-877-1350	Sequence 1350, Ap
26	45.5	10.8	72	9	US-09-738-626-6764	Sequence 6764, Ap
27	45.5	10.8	72	10	US-09-864-761-38891	Sequence 38891, A
28	45	10.6	76	10	US-09-963-896-5	Sequence 5, Appl1
29	44.5	10.5	61	9	US-09-809-391-615	Sequence 615, App
30	44.5	10.5	65	9	US-09-925-299-1416	Sequence 1416, Ap
31	44.5	10.5	65	10	US-09-925-299-1416	Sequence 1416, Ap
32	44	10.4	68	9	US-09-983-802-611	Sequence 611, App
33	44	10.4	68	10	US-09-925-300-1560	Sequence 1560, App
34	44	10.4	79	10	US-09-864-441-460	Sequence 460, App
35	43.5	10.3	52	10	US-09-864-761-44542	Sequence 44542, A
36	43.5	10.3	55	10	US-09-879-957-111	Sequence 111, App
37	43.5	10.3	55	10	US-09-867-550-588	Sequence 588, App
38	43.5	10.3	72	10	US-09-764-877-1500	Sequence 1500, App
39	43	10.2	66	10	US-09-764-877-1799	Sequence 1799, Ap
40	43	10.2	79	9	US-10-002-344A-213	Sequence 213, App
41	42.5	10.0	61	10	US-09-879-957-208	Sequence 208, App
42	42.5	10.0	79	10	US-09-864-761-48228	Sequence 48228, A
43	42.5	10.0	87	10	US-09-864-761-33792	Sequence 33792, A
44	42	9.9	44	10	US-09-864-761-40508	Sequence 40508, A
45	42	9.9	62	10	US-09-864-761-47330	Sequence 47330, A
46	42	9.9	63	10	US-09-864-761-34262	Sequence 34262, A
47	42	9.9	71	9	US-09-986-480-365	Sequence 365, App
48	42	9.9	71	10	US-09-764-860-397	Sequence 397, App
49	42	9.9	9	12	US-10-014-774-9	Sequence 9, Appl1
50	42	9.9	78	10	US-09-795-501-10	Sequence 10, Appl1
51	42	9.9	83	10	US-09-764-869-1093	Sequence 1093, Ap
52	42	9.9	83	10	US-09-864-761-41234	Sequence 41234, Ap
53	41.5	9.8	45	10	US-09-864-761-43127	Sequence 43127, A
54	41.5	9.8	78	9	US-09-909-5678-41	Sequence 41, Appl1
55	41	9.7	43	10	US-09-864-761-42091	Sequence 42091, A
56	41	9.7	53	10	US-09-325-501-1639	Sequence 1639, App
57	41	9.7	63	10	US-09-877-258-2	Sequence 2, Appl1
58	41	9.7	67	9	US-09-796-692-2075	Sequence 2075, Ap
59	41	9.7	67	9	US-09-796-692-2315	Sequence 2315, Ap
60	41	9.7	90	12	US-09-895-913A-280	Sequence 280, App
61	41	9.7	90	10	US-10-001-879-198	Sequence 198, App
62	40.5	9.6	69	10	US-09-982-809-5	Sequence 5, Appl1
63	40.5	9.6	71	9	US-10-001-876-151	Sequence 151, App
64	40	9.5	47	9	US-09-983-802-656	Sequence 656, App
65	40	9.5	49	10	US-09-864-761-41352	Sequence 41352, A
66	40	9.5	56	10	US-09-919-124-48	Sequence 48, Appl1
67	40	9.5	65	10	US-09-864-761-48202	Sequence 48202, A
68	40	9.5	68	9	US-09-984-245-132	Sequence 132, App
69	40	9.5	68	9	US-09-966-262-132	Sequence 132, App
70	40	9.5	69	9	US-10-043-487-398	Sequence 398, App
71	40	9.5	71	9	US-09-764-872-351	Sequence 351, App
72	40	9.5	72	9	US-09-983-802-655	Sequence 655, App
73	40	9.5	72	10	US-09-954-512-1	Sequence 1, Appl1
74	40	9.5	73	10	US-09-864-761-44101	Sequence 44101, A
75	40	9.5	74	9	US-09-986-480-352	Sequence 352, App
76	40	9.5	78	10	US-09-864-761-36451	Sequence 36451, A
77	40	9.5	85	9	US-10-002-344A-167	Sequence 147, App
78	40	9.5	85	10	US-09-864-761-34433	Sequence 34433, A
79	40	9.5	90	10	US-10-001-873-34	Sequence 34, Appl1
80	40	9.5	90	10	US-09-864-761-40114	Sequence 40114, A
81	39.5	9.3	53	9	US-09-796-692-1814	Sequence 1814, Ap
82	39.5	9.3	53	9	US-09-796-692-2071	Sequence 2071, Ap
83	39.5	9.3	68	10	US-09-864-761-44168	Sequence 44168, A
84	39.5	9.3	68	10	US-09-864-761-47104	Sequence 47104, A
85	39.5	9.3	75	10	US-09-867-550-336	Sequence 336, App
86	39.5	9.3	84	10	US-09-867-550-596	Sequence 596, App
87	39.5	9.3	84	10	US-09-925-300-1166	Sequence 1166, Ap
88	39.5	9.3	87	10	US-09-864-761-43703	Sequence 43703, A
89	39.5	9.3	87	10	US-09-864-761-44930	Sequence 44930, A
90	39.5	9.3	89	10	US-09-815-242-4918	Sequence 4918, Ap
91	39	9.2	47	10	US-09-864-761-34601	Sequence 34601, A
92	39	9.2	47	10	US-09-864-761-43815	Sequence 43815, A









531	33.5	7.9	62	10	US-09-764-869-946	Sequence 946, App	604	33	7.8	67	10	US-09-864-761-38332	Sequence 38332, A
532	33.5	7.9	62	10	US-09-764-860-539	Sequence 539, App	605	33	7.8	67	10	US-09-820-893-109	Sequence 109, App
533	33.5	7.9	63	10	US-09-864-761-48173	Sequence 48173, A	606	33	7.8	67	10	US-09-867-550-1694	Sequence 1694, App
534	33.5	7.9	64	10	US-09-864-761-35626	Sequence 35626, A	607	33	7.8	68	10	US-09-795-501-8	Sequence 8, Appl
535	33.5	7.9	65	10	US-10-092-154-852	Sequence 852, App	608	33	7.8	68	10	US-09-864-761-36782	Sequence 36782, A
536	33.5	7.9	65	10	US-09-864-761-39828	Sequence 39828, A	609	33	7.8	68	10	US-10-158-684-16	Sequence 16, Appl
537	33.5	7.9	65	10	US-09-864-761-43927	Sequence 43927, A	610	33	7.8	69	9	US-10-158-711-16	Sequence 16, Appl
538	33.5	7.9	65	10	US-09-764-847-852	Sequence 852, App	611	33	7.8	69	9	US-10-083-337-694	Sequence 694, App
539	33.5	7.9	66	10	US-09-764-847-852	Sequence 852, App	612	33	7.8	69	10	US-09-764-860-377	Sequence 377, App
540	33.5	7.9	66	10	US-09-864-761-45942	Sequence 45942, A	613	33	7.8	69	10	US-09-820-893-52	Sequence 52, Appl
541	33.5	7.9	67	10	US-09-764-864-1082	Sequence 1082, App	614	33	7.8	70	10	US-09-867-550-1774	Sequence 1774, App
542	33.5	7.9	67	10	US-09-738-526-6469	Sequence 6469, App	615	33	7.8	71	9	US-10-077-111-9	Sequence 9, Appl
543	33.5	7.9	73	10	US-09-864-761-36723	Sequence 36723, A	616	33	7.8	71	9	US-10-011-931-31	Sequence 31, Appl
544	33.5	7.9	76	10	US-09-864-761-46901	Sequence 46901, A	617	33	7.8	71	10	US-09-864-761-43211	Sequence 43211, A
545	33.5	7.9	77	9	US-10-016-634A-136	Sequence 136, App	618	33	7.8	72	10	US-09-864-761-43555	Sequence 43555, A
546	33.5	7.9	78	10	US-09-764-870-509	Sequence 509, App	619	33	7.8	72	10	US-09-734-569-111	Sequence 111, App
547	33.5	7.9	78	10	US-09-864-761-35930	Sequence 35930, A	620	33	7.8	72	10	US-09-731-872-251	Sequence 251, App
548	33.5	7.9	80	9	US-09-809-391-471	Sequence 471, App	621	33	7.8	73	9	US-10-063-547-30	Sequence 30, Appl
549	33.5	7.9	80	10	US-09-864-761-35898	Sequence 35898, A	622	33	7.8	73	9	US-10-063-516-30	Sequence 30, Appl
550	33.5	7.9	80	10	US-09-864-761-37250	Sequence 37250, A	623	33	7.8	73	9	US-10-063-502-30	Sequence 30, Appl
551	33.5	7.9	84	9	US-09-864-761-48011	Sequence 48011, A	624	33	7.8	73	9	US-10-006-856A-22	Sequence 22, Appl
552	33.5	7.9	84	9	US-10-178-213-80	Sequence 80, Appl	625	33	7.8	73	9	US-10-063-518-30	Sequence 30, Appl
553	33.5	7.9	85	9	US-09-733-523-2	Sequence 2, Appl	626	33	7.8	73	9	US-10-063-598-30	Sequence 30, Appl
554	33.5	7.9	86	10	US-10-114-093-194	Sequence 194, App	627	33	7.8	73	9	US-10-006-893A-22	Sequence 22, Appl
555	33.5	7.9	86	10	US-09-864-761-34118	Sequence 34118, A	628	33	7.8	73	9	US-10-063-518-30	Sequence 30, Appl
556	33.5	7.9	88	10	US-09-764-860-1057	Sequence 1057, App	629	33	7.8	73	9	US-10-063-518-30	Sequence 30, Appl
557	33.5	7.9	88	10	US-09-925-300-1321	Sequence 1321, App	630	33	7.8	73	12	US-10-006-867-30	Sequence 30, Appl
558	33.5	7.9	89	9	US-09-843-676-204	Sequence 204, App	631	33	7.8	74	9	US-09-796-692-1973	Sequence 1973, App
559	33.5	7.9	89	9	US-09-438-786-204	Sequence 204, App	632	33	7.8	74	9	US-09-796-692-2489	Sequence 2489, App
560	33.5	7.9	89	9	US-10-053-758-204	Sequence 204, App	633	33	7.8	74	10	US-09-898-416-17	Sequence 17, Appl
561	33	7.9	89	9	US-10-054-295-204	Sequence 204, App	634	33	7.8	76	10	US-10-092-154-978	Sequence 154, Appl
562	33	7.8	40	10	US-09-864-761-34563	Sequence 34563, A	635	33	7.8	76	10	US-09-764-887-167	Sequence 167, App
563	33	7.8	41	10	US-09-864-761-43796	Sequence 43796, A	636	33	7.8	76	10	US-09-764-847-978	Sequence 978, App
564	33	7.8	44	10	US-09-864-761-33432	Sequence 33432, A	637	33	7.8	77	9	US-09-796-692-1417	Sequence 1417, App
565	33	7.8	44	10	US-09-864-761-43450	Sequence 43450, A	638	33	7.8	77	9	US-09-796-692-1885	Sequence 1885, App
566	33	7.8	46	9	US-09-939-980-274	Sequence 274, App	639	33	7.8	77	10	US-09-864-761-39700	Sequence 39700, A
567	33	7.8	46	9	US-09-813-153-158	Sequence 158, App	640	33	7.8	77	10	US-09-867-550-1256	Sequence 1256, App
568	33	7.8	46	9	US-09-948-820-80	Sequence 80, Appl	641	33	7.8	78	9	US-10-002-344A-206	Sequence 206, App
569	33	7.8	46	10	US-09-864-761-44538	Sequence 44538, A	642	33	7.8	78	9	US-10-002-344A-207	Sequence 207, App
570	33	7.8	49	10	US-10-112-455A-16	Sequence 16, Appl	643	33	7.8	78	9	US-09-738-626-5645	Sequence 5645, App
571	33	7.8	50	9	US-09-864-761-47113	Sequence 47113, A	644	33	7.8	81	9	US-09-984-245-165	Sequence 165, App
572	33	7.8	50	9	US-10-190-438-19	Sequence 19, Appl	645	33	7.8	81	9	US-09-796-692-2466	Sequence 2466, App
573	33	7.8	50	10	US-10-190-438-20	Sequence 20, Appl	646	33	7.8	81	9	US-09-966-262-165	Sequence 165, App
574	33	7.8	52	10	US-09-864-761-38825	Sequence 38825, A	647	33	7.8	81	10	US-09-864-761-44708	Sequence 44708, A
575	33	7.8	52	10	US-09-864-761-44474	Sequence 44474, A	648	33	7.8	81	10	US-09-867-550-210	Sequence 210, App
576	33	7.8	54	10	US-09-867-550-290	Sequence 290, App	649	33	7.8	82	10	US-09-764-877-1894	Sequence 1894, App
577	33	7.8	54	10	US-09-864-761-47986	Sequence 47986, A	650	33	7.8	84	9	US-10-092-154-592	Sequence 592, App
578	33	7.8	54	10	US-09-867-550-1586	Sequence 1586, App	651	33	7.8	84	10	US-09-864-761-39755	Sequence 39755, A
579	33	7.8	55	9	US-09-764-877-1820	Sequence 1820, App	652	33	7.8	84	10	US-09-864-761-39807	Sequence 39807, A
580	33	7.8	55	9	US-09-158-722-2	Sequence 2, Appl	653	33	7.8	84	10	US-09-764-847-592	Sequence 458, App
581	33	7.8	57	9	US-09-864-761-38194	Sequence 38194, A	654	33	7.8	85	10	US-09-764-853-458	Sequence 1668, App
582	33	7.8	57	10	US-09-983-802-534	Sequence 534, App	655	33	7.8	88	9	US-10-083-357-971	Sequence 971, Appl
583	33	7.8	58	10	US-09-864-761-47845	Sequence 47845, A	656	33	7.8	88	9	US-09-950-933A-56	Sequence 56, Appl
584	33	7.8	58	10	US-09-864-761-35718	Sequence 35718, A	657	33	7.8	42	10	US-09-764-869-888	Sequence 888, App
585	33	7.8	60	10	US-09-864-761-34914	Sequence 34914, A	658	32.5	7.7	46	10	US-09-864-761-37079	Sequence 37079, A
586	33	7.8	60	10	US-09-864-761-46375	Sequence 46375, A	659	32.5	7.7	46	10	US-09-864-761-42434	Sequence 42434, A
587	33	7.8	62	9	US-09-867-550-520	Sequence 520, App	660	32.5	7.7	49	9	US-09-982-802-241	Sequence 241, App
588	33	7.8	62	10	US-10-050-704-334	Sequence 334, App	661	32.5	7.7	49	10	US-09-864-761-42434	Sequence 42434, A
589	33	7.8	62	10	US-09-864-761-48027	Sequence 48027, A	662	32.5	7.7	50	9	US-09-834-133-733	Sequence 733, App
590	33	7.8	62	10	US-09-820-893-91	Sequence 91, Appl	663	32.5	7.7	50	10	US-09-220-920-108	Sequence 108, App
591	33	7.8	62	10	US-09-967-552A-22	Sequence 22, Appl	664	32.5	7.7	51	10	US-09-071-838-167	Sequence 167, App
592	33	7.8	62	10	US-09-967-552A-58	Sequence 58, Appl	665	32.5	7.7	51	10	US-09-864-761-46322	Sequence 46322, A
593	33	7.8	63	10	US-09-864-761-42311	Sequence 42311, A	666	32.5	7.7	51	10	US-09-864-761-33438	Sequence 33438, A
594	33	7.8	63	10	US-09-864-761-42311	Sequence 42311, A	667	32.5	7.7	52	10	US-09-864-761-35037	Sequence 35037, A
595	33	7.8	64	9	US-09-820-893-79	Sequence 79, Appl	668	32.5	7.7	52	10	US-09-904-621-8	Sequence 8, Appl
596	33	7.8	64	9	US-09-796-692-1638	Sequence 1638, App	669	32.5	7.7	54	10	US-09-764-898-273	Sequence 273, App
597	33	7.8	64	10	US-09-864-761-1378	Sequence 1378, App	670	32.5	7.7	54	10	US-09-925-301-1162	Sequence 1162, App
598	33	7.8	65	9	US-09-864-761-34993	Sequence 34993, A	671	32.5	7.7	58	10	US-09-764-877-1406	Sequence 1406, App
599	33	7.8	65	10	US-09-864-761-39839	Sequence 39839, A	672	32.5	7.7	59	10	US-09-864-761-42204	Sequence 42204, A
600	33	7.8	66	10	US-09-925-299-1037	Sequence 1037, App	673	32.5	7.7	59	10	US-09-867-550-1524	Sequence 1524, App
601	33	7.8	66	10	US-09-864-761-46272	Sequence 46272, A	674	32.5	7.7	60	10	US-09-843-598-8	Sequence 8, Appl
602	33	7.8	66	10	US-09-867-550-728	Sequence 728, App	675	32.5	7.7	61	9	US-10-001-835-192	Sequence 192, App
603	33	7.8	67	9	US-09-967-552A-60	Sequence 60, Appl	676	32.5	7.7	61	9	US-09-981-876-138	Sequence 138, App







```

1 / PRIOR FILING DATE: 2001-01-17
2 / PRIOR APPLICATION NUMBER: 09/764,892
3 / PRIOR FILING DATE: 2001-01-17
4 / PRIOR APPLICATION NUMBER: US01/01345
5 / PRIOR FILING DATE: 2001-01-17
6 / PRIOR APPLICATION NUMBER: 09/764,888
7 / PRIOR FILING DATE: 2001-01-17
8 / PRIOR APPLICATION NUMBER: US01/01329
9 / PRIOR FILING DATE: 2001-01-17
10 / PRIOR APPLICATION NUMBER: 09/764,905
11 / PRIOR FILING DATE: 2001-01-17
12 / PRIOR APPLICATION NUMBER: US01/01354
13 / PRIOR FILING DATE: 2001-01-17
14 / PRIOR APPLICATION NUMBER: 09/764,891
15 / PRIOR FILING DATE: 2001-01-17
16 / PRIOR APPLICATION NUMBER: US01/01339
17 / PRIOR FILING DATE: 2001-01-17
18 / PRIOR APPLICATION NUMBER: 09/764,869
19 / PRIOR FILING DATE: 2001-01-17
20 / PRIOR APPLICATION NUMBER: US01/01340
21 / PRIOR FILING DATE: 2001-01-17
22 / PRIOR APPLICATION NUMBER: 09/764,874
23 / PRIOR FILING DATE: 2001-01-17
24 / PRIOR APPLICATION NUMBER: US01/01334
25 / PRIOR FILING DATE: 2001-01-17
26 / PRIOR APPLICATION NUMBER: 09/764,898
27 / PRIOR FILING DATE: 2001-01-17
28 / PRIOR APPLICATION NUMBER: US01/01320
29 / PRIOR FILING DATE: 2001-01-17
30 / PRIOR APPLICATION NUMBER: 09/764,853
31 / PRIOR FILING DATE: 2001-01-17
32 / PRIOR APPLICATION NUMBER: US01/01349
33 / PRIOR FILING DATE: 2001-01-17
34 / PRIOR APPLICATION NUMBER: 09/764,902
35 / PRIOR FILING DATE: 2001-01-17
36 / PRIOR APPLICATION NUMBER: US01/01239
37 / PRIOR FILING DATE: 2001-01-17
38 / PRIOR APPLICATION NUMBER: 09/764,870
39 / PRIOR FILING DATE: 2001-01-17
40 / PRIOR APPLICATION NUMBER: US01/01348
41 / PRIOR FILING DATE: 2001-01-17
42 / PRIOR APPLICATION NUMBER: 09/764,882
43 / PRIOR FILING DATE: 2001-01-17
44 / PRIOR APPLICATION NUMBER: US01/01347
45 / PRIOR FILING DATE: 2001-01-17
46 / PRIOR APPLICATION NUMBER: 09/764,896
47 / PRIOR FILING DATE: 2001-01-17
48 / PRIOR APPLICATION NUMBER: US01/01307
49 / PRIOR FILING DATE: 2001-01-17
50 / PRIOR APPLICATION NUMBER: 09/764,864
51 / PRIOR FILING DATE: 2001-01-17
52 / PRIOR APPLICATION NUMBER: US01/01341
53 / PRIOR FILING DATE: 2001-01-17
54 / PRIOR APPLICATION NUMBER: 09/764,856
55 / PRIOR FILING DATE: 2001-01-17
56 / PRIOR APPLICATION NUMBER: US01/01336
57 / PRIOR FILING DATE: 2001-01-17
58 / PRIOR APPLICATION NUMBER: 09/764,868
59 / PRIOR FILING DATE: 2001-01-17
60 / PRIOR APPLICATION NUMBER: US01/01312
61 / PRIOR FILING DATE: 2001-01-17
62 / PRIOR APPLICATION NUMBER: 60/179,065
63 / PRIOR FILING DATE: 2000-01-31
64 / PRIOR APPLICATION NUMBER: 60/180,628
65 / PRIOR FILING DATE: 2000-02-04
66 / PRIOR APPLICATION NUMBER: 60/209,467
67 / PRIOR FILING DATE: 2000-06-07
68 / NUMBER OF SEQ ID NOS: 167
69 / SOFTWARE: PatentIn Ver. 2.0
70 / SEQ ID NO 103
71 / LENGTH: 83
72 / TYPE: PRT
73 / ORGANISM: Homo sapiens

```

```

; FEATURE:
; NAME/REV: SITE
; LOCATION: (25)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-908-711-103

Query Match      12.2%   Score 51.5; DB 10; Length 83;
Best Local Similarity 27.5%; Pred. No. 41;
Matches 19; Conservative 8; Mismatches 33; Indels 9; Gaps 2;

QY 15 RSISENSLVAMDF-SGQKSRVIENTEALSVAVEEGIAWRKK-----GCLRLGTGHS 65
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 9 RDVGDDLPQMEVSGXGXGRPRTPPASGPRIHSSRRKAPWRRRLPQWNPNCGTRPGSAAQ 68
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

QY 66 PTASSQSQA 74
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 69 PWGSSOASS 77
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

RESULT 4
US-09-864-761-47521
; Sequence 47521, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL
; FILE REFERENCE: Aecomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 47521
; LENGTH: 84

```



Query Match 10.9%; Score 46; DB 9; Length 68;  
Best Local Similarity 47.4%; Pred. No. 1.4e+02;

```

RESULT 12
US-10-092-154-700
: Sequence 700, Application US/10092154
: Publication No. US20030054375A1
: GENERAL INFORMATION:
: APPLICANT: Rosen et al.
: TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
: FILE REFERENCE: PC009C1
: CURRENT APPLICATION NUMBER: US/10/092,154
: CURRENT FILING DATE: 2002-03-07

```

```
; NUMBER OF SEQ ID NOS: 2003
; Prior Application removed - See File Wrapper or Palm
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 700
; LENGTH: 69
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (57)
; OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids
US-10-092-154-700

Query Match          10.9%; Score 46; DB 9; Length 69;
Best Local Similarity 37.9%; Pred. No. 1.5e+02;
Matches 11; Conservative 3; Mismatches 15; Indels 0; Gaps 0;

QY 56 GCLRLGTHGSPPTASSOSATNMAIHRSP 84
DB 31 GLKLLGSSPPPTLASOSAGITGMHCXOP 59

RESULT 13
US-09-764-847-700
; Sequence 700, Application US/09764847
; Patent No. US20020132767A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC009
; CURRENT APPLICATION NUMBER: US/09/764,847
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 2003
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 700
; LENGTH: 69
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (57)
; OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids
US-09-764-847-700

Query Match          10.9%; Score 46; DB 10; Length 69;
Best Local Similarity 37.9%; Pred. No. 1.5e+02;
Matches 11; Conservative 3; Mismatches 15; Indels 0; Gaps 0;

QY 56 GCLRLGTHGSPPTASSOSATNMAIHRSP 84
DB 31 GLKLLGSSPPPTLASOSAGITGMHCXOP 59

RESULT 14
US-09-925-299-1460
; Sequence 1460, Application US/09925299
; Publication No. US20030040617A9
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299
; Prior Application Number: 60/124,270
; Prior Filing Date: 2001-08-10
; Prior Application Number: 60/124,270
; Prior Filing Date: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1460
; LENGTH: 56
; TYPE: PRT
```

```
; ORGANISM: Homo sapiens
US-09-925-299-1460

Query Match          10.8%; Score 45.5; DB 9; Length 56;
Best Local Similarity 35.4%; Pred. No. 1.3e+02;
Matches 17; Conservative 6; Mismatches 16; Indels 9; Gaps 3;

QY 37 NPTTEALSVAVE--EGLAWRKKGCG---LRLGTHGSPPTASSO-----SSAT 75
DB 6 NPLVNLTVSPKRNSSLDTRKKPCRESKKFNTHSRPKSSHOLRKRSST 53

RESULT 15
US-09-925-299-1460
; Sequence 1460, Application US/09925299
; Patent No. US20020055627A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299
; Prior Application Number: PCT/US00/05883
; Prior Filing Date: 2001-08-10
; Prior Application Number: 60/124,270
; Prior Filing Date: 2000-03-08
; Prior Application Number: 60/124,270
; Prior Filing Date: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1460
; LENGTH: 56
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-925-299-1460

Query Match          10.8%; Score 45.5; DB 10; Length 56;
Best Local Similarity 35.4%; Pred. No. 1.3e+02;
Matches 17; Conservative 6; Mismatches 16; Indels 9; Gaps 3;

QY 37 NPTTEALSVAVE--EGLAWRKKGCG---LRLGTHGSPPTASSO-----SSAT 75
DB 6 NPLVNLTVSPKRNSSLDTRKKPCRESKKFNTHSRPKSSHOLRKRSST 53

RESULT 16
US-09-925-299-1372
; Sequence 1372, Application US/09925299
; Publication No. US20030040617A9
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299
; Prior Application Number: 60/124,270
; Prior Filing Date: 2001-08-10
; Prior Application Number: 60/124,270
; Prior Filing Date: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1372
; LENGTH: 58
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-925-299-1372

Query Match          10.8%; Score 45.5; DB 9; Length 58;
Best Local Similarity 35.4%; Pred. No. 1.3e+02;
Matches 17; Conservative 6; Mismatches 16; Indels 9; Gaps 3;

QY 37 NPTTEALSVAVE--EGLAWRKKGCG---LRLGTHGSPPTASSO-----SSAT 75
DB 8 NPLVNLTVSPKRNSSLDTRKKPCRESKKFNTHSRPKSSHOLRKRSST 55
```





```

; LENGTH: 58
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-925-299-1372

Query Match      10.8%; Score 45.5; DB 10; Length 58;
Best Local Similarity 35.4%; Pred. No. 1.3e+02;
Matches 17; Conservative 6; Mismatches 16; Indels 9; Gaps 3;

QY 37 NPTEALSVAVE--EGLAWRKKG---LRIGTHGSPASSQ-----SSAT 75
Db 8 NPLVNLTVSPKRNSSLDTRKKPCRESKFKFNTHSRPKSSHOLKRKSSST 55

RESULT 22
US-09-925-299-1437
; Sequence 1437, Application US/09925299
; Patent No. US20020055627A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1437
; LENGTH: 58
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-925-299-1437

Query Match      10.8%; Score 45.5; DB 10; Length 58;
Best Local Similarity 35.4%; Pred. No. 1.3e+02;
Matches 17; Conservative 6; Mismatches 16; Indels 9; Gaps 3;

QY 37 NPTEALSVAVE--EGLAWRKKG---LRIGTHGSPASSQ-----SSAT 75
Db 8 NPLVNLTVSPKRNSSLDTRKKPCRESKFKFNTHSRPKSSHOLKRKSSST 55

RESULT 23
US-09-925-299-1493
; Sequence 1493, Application US/09925299
; Patent No. US20020055627A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1493
; LENGTH: 58
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (4)
; OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids
US-09-925-299-1493

Query Match      10.8%; Score 45.5; DB 10; Length 58;
Best Local Similarity 35.4%; Pred. No. 1.3e+02;

```

```

Matches 17; Conservative 6; Mismatches 16; Indels 9; Gaps 3;

QY 37 NPTEALSVAVE--EGLAWRKKG---LRIGTHGSPASSQ-----SSAT 75
Db 8 NPLVNLTVSPKRNSSLDTRKKPCRESKFKFNTHSRPKSSHOLKRKSSST 55

RESULT 24
US-09-925-299-1528
; Sequence 1528, Application US/09925299
; Patent No. US20020055627A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1528
; LENGTH: 58
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-925-299-1528

Query Match      10.8%; Score 45.5; DB 10; Length 58;
Best Local Similarity 35.4%; Pred. No. 1.3e+02;
Matches 17; Conservative 6; Mismatches 16; Indels 9; Gaps 3;

QY 37 NPTEALSVAVE--EGLAWRKKG---LRIGTHGSPASSQ-----SSAT 75
Db 8 NPLVNLTVSPKRNSSLDTRKKPCRESKFKFNTHSRPKSSHOLKRKSSST 55

RESULT 25
US-09-764-877-1350
; Sequence 1350, Application US/09764877
; Patent No. US20020147140A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC005
; CURRENT APPLICATION NUMBER: US/09/764,877
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 4031
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1350
; LENGTH: 60
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-877-1350

Query Match      10.8%; Score 45.5; DB 10; Length 60;
Best Local Similarity 33.3%; Pred. No. 1.4e+02;
Matches 16; Conservative 6; Mismatches 25; Indels 1; Gaps 1;

QY 38 PTEALSVAVEEGLAWRKKGCLRLGTHGS-PTASSQSATNNATHRSQP 84
Db 6 PCQFFVFLVEMGFHHVQCAGLELLTSGDPTTSASQSAGITGMSHCAQP 53

Search completed: March 28, 2003, 09:19:20
Job time : 31.1181 secs

```

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 28, 2003, 09:03:25 ; Search time 24.4724 Seconds  
(without alignments)  
100.992 Million cell updates/sec

Sequence: 1 QGRSGSSOSISPMRISSEN.....SPTASSQSATNMAIHRSQP 84

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 20297

Minimum DB seq length: 40

Maximum DB seq length: 90

Post-processing: Minimum Match 0%

Listing first 1000 summaries

- 1: /cgn2.6/ptodata/1/iaa/5A\_COMB.pep.\*
- 2: /cgn2.6/ptodata/1/iaa/5B\_COMB.pep.\*
- 3: /cgn2.6/ptodata/1/iaa/6A\_COMB.pep.\*
- 4: /cgn2.6/ptodata/1/iaa/6B\_COMB.pep.\*
- 5: /cgn2.6/ptodata/1/iaa/PCTUS\_COMB.pep.\*
- 6: /cgn2.6/ptodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	50	11.8	68	3	US-09-100-804-30
2	45.5	10.8	73	4	US-09-331-930A-22
3	45	10.6	76	3	US-09-083-521-5
4	44.5	10.5	61	4	US-09-349-476-615
5	44	10.4	47	4	US-08-776-059-18
6	44	10.4	63	4	US-09-227-357-611
7	44	10.4	80	3	US-09-081-320-20
8	44	10.4	80	4	US-09-574-141A-20
9	44	10.4	80	4	US-09-707-780-20
10	43.5	10.3	55	4	US-08-630-915A-111
11	43.5	10.3	74	4	US-09-267-177-12
12	43	10.2	53	2	US-08-726-306A-144
13	43	10.2	84	4	US-09-247-155-173
14	42.5	10.0	61	4	US-08-630-915A-208
15	42	9.9	40	1	US-07-641-971B-5
16	42	9.9	40	1	US-07-781-248A-5
17	42	9.9	47	4	US-08-776-059-16
18	42	9.9	78	4	US-09-134-001C-2848
19	42	9.9	84	6	5171684-3
20	41.5	9.8	87	3	US-09-023-082A-76
21	41	9.7	51	4	US-08-927-219-49
22	41	9.7	76	4	US-09-246-500B-9
23	40.5	9.6	66	2	US-08-459-568-52
24	40.5	9.6	66	2	US-08-399-411-52
25	40.5	9.6	66	3	US-08-516-859A-52
26	40.5	9.6	66	4	US-09-586-472-52
27	40.5	9.6	66	4	US-09-528-706-52















Sequence 11, Appl  
Sequence 4, Appl  
Sequence 12, Appl  
Sequence 12, Appl  
Sequence 12, Appl  
Sequence 12, Appl  
Sequence 6, Appl  
Sequence 15, Appl  
Sequence 6, Appl  
Sequence 427, Appl  
Sequence 552, Appl  
Sequence 4, Appl  
Sequence 6, Appl  
Sequence 3, Appl  
Sequence 99, Appl  
Sequence 199, Appl  
Sequence 15, Appl  
Sequence 27, Appl  
Sequence 71, Appl  
Sequence 71, Appl  
Sequence 71, Appl  
Sequence 4, Appl  
Sequence 23, Appl

977 29.5 7.0 90 4 US-09-199-534-11  
978 29 6.9 40 1 US-08-272-882b-4  
979 29 6.9 40 1 US-08-188-228-12  
980 29 6.9 40 1 US-08-332-643-12  
981 29 6.9 40 1 US-08-332-638-12  
982 29 6.9 40 2 US-08-679-405-6  
983 29 6.9 40 2 US-08-797-842-15  
984 29 6.9 40 2 US-08-842-799-6  
985 29 6.9 40 4 US-08-936-165A-427  
986 29 6.9 40 4 US-08-469-260A-552  
987 29 6.9 40 5 PCT-US92-05532-4  
988 29 6.9 40 5 PCT-US96-11458-6  
989 29 6.9 41 1 US-07-700-526-3  
990 29 6.9 41 5 PCT-US92-03132-3  
991 29 6.9 42 1 US-08-282-037-99  
992 29 6.9 42 4 US-08-282-773-199  
993 29 6.9 42 4 US-08-942-012B-15  
994 29 6.9 42 4 US-09-131-750-27  
995 29 6.9 43 1 US-07-998-003A-71  
996 29 6.9 43 1 US-08-453-274B-71  
997 29 6.9 43 1 US-08-453-695A-71  
998 29 6.9 43 1 US-08-268-161A-71  
999 29 6.9 43 2 US-08-599-602-4  
1000 29 6.9 43 2 US-08-488-161-23

## ALIGNMENTS

RESULT 1  
US-09-100-804-30  
Sequence 30, Application US/09100804  
Patent No. 6066472  
GENERAL INFORMATION:  
APPLICANT: GONZALEZ, LEONEL JORGE  
APPLICANT: SARAS, JAN  
APPLICANT: CLAESSON-WELSH, LENA  
APPLICANT: HELDIN, CARL-HENRIK  
TITLE OF INVENTION: PRIMARY STRUCTURE AND FUNCTIONAL  
TITLE OF INVENTION: EXPRESSION OF NUCLEOTIDE SEQUENCES FOR NOVEL PROTEIN  
TITLE OF INVENTION: TYROSINE PHOSPHATASES  
NUMBER OF SEQUENCES: 34  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.  
STREET: 600 ATLANTIC AVENUE  
CITY: BOSTON  
STATE: MASSACHUSETTS  
COUNTRY: USA  
ZIP: 02210  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentlip Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/100,804  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/596,291  
FILING DATE: 09-AUG-1996  
APPLICATION NUMBER: US 08/115,573  
FILING DATE: 01-SEP-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US94/09943  
FILING DATE: 01-SEP-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: GATES, EDWARD R.  
REGISTRATION NUMBER: 31,616  
REFERENCE/DOCKET NUMBER: L0461/7003  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-720-3500  
TELEFAX: 617-720-2441

TELEX:  
INFORMATION FOR SEQ ID NO: 30:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 68 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: Peptide  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-09-100-804-30  
Query Match 11.8%; Score 50; DB 3; Length 68;  
Best Local Similarity 31.4%; Pred. No. 17;  
Matches 16; Conservative 11; Mismatches 20; Indels 4; Gaps 1;  
QY 14 MRSISENSLVAMDFSGQSRVI----ENPTALSVAVEEGLAWRKGCLRL 60  
DB 16 VKAISQDSLAARDGDIQEGDVVKINGTVTENMSLTDAKTLIERSKGKLM 66  
RESULT 2  
US-09-331-930A-22  
Sequence 22, Application US/09331930A  
Patent No. 6436670  
GENERAL INFORMATION:  
APPLICANT: ZIMMET, PAUL Z.  
APPLICANT: COLLIER, GREGORY  
TITLE OF INVENTION: A NOVEL GENE AND USES THEREFOR  
FILE REFERENCE: 22975-20007 00  
CURRENT APPLICATION NUMBER: US/09/331,930A  
CURRENT FILING DATE: 1999-06-30  
PRIOR APPLICATION NUMBER: PCI/AU98/00902  
PRIOR FILING DATE: 1998-10-30  
PRIOR APPLICATION NUMBER: AU PP0117/97  
PRIOR FILING DATE: 1997-10-31  
PRIOR APPLICATION NUMBER: AU PP0323/97  
PRIOR FILING DATE: 1997-11-11  
NUMBER OF SEQ ID NOS: 27  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 22  
LENGTH: 73  
TYPE: PRT  
ORGANISM: Caenorhabditis elegans  
US-09-331-930A-22  
Query Match 10.8%; Score 45.5; DB 4; Length 73;  
Best Local Similarity 29.4%; Pred. No. 77;  
Matches 10; Conservative 7; Mismatches 12; Indels 5; Gaps 1;  
QY 26 DFGSGKSRVTEALSTALS-----VAVEEGLAWRK 54  
DB 8 DLGKKVKIKCNFSFTIGDLKKLIAAQTGTREK 41  
RESULT 3  
US-09-083-521-5  
Sequence 5, Application US/09083521  
Patent No. 6048970  
GENERAL INFORMATION:  
APPLICANT: Lal, Preeti  
APPLICANT: Guegler, Karl J.  
APPLICANT: Corley, Neil C.  
TITLE OF INVENTION: PROSTATE GROWTH-ASSOCIATED MEMBRANE PROTEINS  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.  
STREET: 3174 PORTER DRIVE  
CITY: PALO ALTO  
STATE: CALIFORNIA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:

```

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/083,521
; FILING DATE: Herewith
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: CERRONE, MICHAEL C.
; REGISTRATION NUMBER: 39,132
; REFERENCE/DOCKET NUMBER: PF-0527 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 855-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 76 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 1216498
; US-09-083-521-5

```

```

Query Match 10.6%; Score 45; DB 3; Length 76;
Best Local Similarity 24.4%; Pred. No. 95;
Matches 19; Conservative 10; Mismatches 15; Indels 34; Gaps 4;

QY 6 CSSQISPMRSISNSLVAMDSGOKS-RVTENTPEALSVAVEGLAWKKCCLRLGTHG 64
Db 1:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
26 CQNTSVAP-----FSGNQISAAPNPNTNATP-----RSGC----- 55

QY 65 SPTASSQSSATNMAIHS 82
Db 1:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
56 ----SSLOSTAGLLALSLS 70

```

```

RESULT 4
US-09-149-476-615
; Sequence 615, Application US/09149476
; Patent No. 6420526
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 186 Human Secreted proteins
; FILE REFERENCE: P2002P1
; CURRENT APPLICATION NUMBER: US/09/149,476
; EARLIER FILING DATE: 1998-09-08
; EARLIER APPLICATION NUMBER: PCT/US98/04493
; EARLIER FILING DATE: 1998-03-06
; EARLIER APPLICATION NUMBER: 60/040,162
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,333
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/038,621
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,626
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,334
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,336
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,163
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/047,600
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,615
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,597
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,502
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,633
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,583
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,617
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,618
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,503
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,592
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,581
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,584
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,500
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,587
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,492
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,598
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,613
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,582
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,596
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,612
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,632
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,601
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/043,580
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,568
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,314
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,569
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,311
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,671
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,674
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,669
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,312
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,313
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,672
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/043,315
; EARLIER FILING DATE: 1997-04-11
; EARLIER APPLICATION NUMBER: 60/048,974
; EARLIER FILING DATE: 1997-06-06
; EARLIER APPLICATION NUMBER: 60/056,886
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,877
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,889
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,893
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,630
; EARLIER FILING DATE: 1997-08-22
; EARLIER APPLICATION NUMBER: 60/056,878

```

```

/ EARLIER APPLICATION NUMBER: 60/056,900
/ EARLIER FILING DATE: 1997-08-22
/ EARLIER APPLICATION NUMBER: 60/056,875
/ EARLIER FILING DATE: 1997-08-22
/ EARLIER APPLICATION NUMBER: 60/056,862
/ EARLIER FILING DATE: 1997-08-22
/ EARLIER APPLICATION NUMBER: 60/056,887
/ EARLIER FILING DATE: 1997-08-22
/ EARLIER APPLICATION NUMBER: 60/056,908
/ EARLIER FILING DATE: 1997-08-22
/ EARLIER APPLICATION NUMBER: 60/048,964
/ EARLIER FILING DATE: 1997-06-06
/ EARLIER APPLICATION NUMBER: 60/057,650
/ EARLIER FILING DATE: 1997-09-05
/ EARLIER APPLICATION NUMBER: 60/056,884
/ EARLIER FILING DATE: 1997-08-22
/ EARLIER APPLICATION NUMBER: 60/057,669
/ EARLIER FILING DATE: 1997-09-05
/ EARLIER APPLICATION NUMBER: 60/049,610
/ EARLIER FILING DATE: 1997-06-13
/ EARLIER APPLICATION NUMBER: 60/061,060
/ EARLIER FILING DATE: 1997-10-02

Query Match 10.5%; Score 44
Best Local Similarity 37.9%; Pred. NO
Matches 11; Conservative 7; Mismatch

QY 51 AWRKKGCLRGTHGSPASSQSSANTMAI 75
||| | | | | | | | | | | | | | |
Db 33 AWRPSG-----GTGTFSSQSSQTQSR 54

RESULT 5
US-08-776-059-18
/ Sequence 18, Application US/08776059B
/ Patent No. 6271368
/ GENERAL INFORMATION:
/ APPLICANT: LENTZEN, Hans
/ APPLICANT: ECK, Jurgen
/ APPLICANT: BAUR, Axel
/ APPLICANT: ZINKE, Holger
/ TITLE OF INVENTION: RECOMBINANT MISPLETEIN
/ FILE REFERENCE: 674503-2003
/ CURRENT APPLICATION NUMBER: US/08/7776,6
/ CURRENT FILING DATE: 1999-06-19
/ EARLIER APPLICATION NUMBER: PCT/EP96/02
/ EARLIER FILING DATE: 1996-06-25
/ EARLIER APPLICATION NUMBER: 95109949.8
/ EARLIER FILING DATE: 1995-06-26
/ NUMBER OF SEQ ID NOS: 56
/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO 18
/ LENGTH: 47
/ TYPE: PPT
/ ORGANISM: Saponaria officinalis
US-08-776-059-18

Query Match 10.4%; Score 44
Best Local Similarity 34.8%; Pred. NO
Matches 8; Conservative 7; Mismatch

QY 25 MDFSQKSRVIENTEALSVAVE 47
|| | | | | | | | | | | | | |
Db 6 MDAVNKAQVWKNAREFLIAIQ 28

RESULT 6
US-09-227-357-611
/ Sequence 611, Application US/09227357
/ Patent No. 6342581
/ GENERAL INFORMATION:
/ APPLICANT: Fischer et al.
/ TITLE OF INVENTION: 123 Human Secreted

```

Best Local Similarity 37.9%; Pred. No. 80;  
Matches 11; Conservative 4; Mismatches 7; Indels 7; Gaps 1;

## RESULT 5

Query Match . 10.4%; Score 44; DB 4; Length 47;  
Best Local Similarity 34.8%; Pred. No. 64;  
Matches 8; Conservative 7; Mismatches 8; Indels

## RESULT 6

RECD: US-09-227-357-611  
 ; Sequence 611, Application US/09227357  
 ; Patent No. 6342581  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Fischer et al.  
 ; TITLE OF INVENTION: 123 Human Secreted Proteins

[illegible]

```

RESULT 8
US-09-574-141A-20
; Sequence 20, Application US/09574141A
; Patent No. 6395490
; GENERAL INFORMATION:
; APPLICANT: Gonsalves, Dennis
; APPLICANT: Meng, Baozhong
; TITLE OF INVENTION: RUPESTRIS STEM PITTING ASSOCIATED VIRUS
; FILE OF INVENTION: NUCLEIC ACIDS, PROTEINS, AND THEIR USES
; FILE REFERENCE: 07678/035005
; CURRENT APPLICATION NUMBER: US/09/574,141A
; CURRENT FILING DATE: 2000-05-18
; PRIOR FILING DATE: 1997-05-20
; PRIOR APPLICATION NUMBER: 60/047,147
; PRIOR FILING DATE: 1997-05-20
; PRIOR APPLICATION NUMBER: 60/069,902
; PRIOR FILING DATE: 1997-12-17
; PRIOR APPLICATION NUMBER: 09/081,320
; PRIOR FILING DATE: 1998-05-19
; NUMBER OF SEQ ID NOS: 97
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 20
; LENGTH: 80
; TYPE: PRT
; ORGANISM: Rupestris stem pitting associated virus
US-09-574-141A-20

Query Match          10.4%; Score 44; DB 4; Length 80;
Best Local Similarity 50.0%; Pred. No. 1.4e+02;
Matches 13; Conservative 4; Mismatches 7; Indels 2; Gaps 1;

QY 31 KSRVLEN--PTEALSVAVEGLAWRK 54
Db 40 ESIIVNGPSEALATVKEVLGGLK 65

RESULT 9
US-09-707-780-20
; Sequence 20, Application US/09707780
; Patent No. 6399308
; GENERAL INFORMATION:
; APPLICANT: Gonsalves, Dennis
; APPLICANT: Meng, Baozhong
; TITLE OF INVENTION: RUPESTRIS STEM PITTING ASSOCIATED VIRUS
; FILE OF INVENTION: NUCLEIC ACIDS, PROTEINS, AND THEIR USES
; FILE REFERENCE: 07678/035006
; CURRENT APPLICATION NUMBER: US/09/707,780
; CURRENT FILING DATE: 2000-11-07
; PRIOR FILING DATE: 1998-05-19
; PRIOR APPLICATION NUMBER: 60/047,147
; PRIOR FILING DATE: 1997-05-20
; PRIOR APPLICATION NUMBER: 60/069,902
; PRIOR FILING DATE: 1997-12-17
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 20
; LENGTH: 80
; TYPE: PRT
; ORGANISM: Rupestris stem pitting associated virus
US-09-707-780-20

Query Match          10.4%; Score 44; DB 4; Length 80;
Best Local Similarity 50.0%; Pred. No. 1.4e+02;
Matches 13; Conservative 4; Mismatches 7; Indels 2; Gaps 1;

QY 31 KSRVLEN--PTEALSVAVEGLAWRK 54
Db 40 ESIIVNGPSEALATVKEVLGGLK 65

RESULT 10
US-08-630-915A-111
; Sequence 11, Application US/08630915A
; Patent No. 6309820
; GENERAL INFORMATION:
; APPLICANT: SPARKS, Andrew B.
; APPLICANT: HOFFMAN, No. 6309820h
; APPLICANT: KAY, Brian K.
; APPLICANT: FOMLKE, Dana M.
; APPLICANT: MCCONNELL, Stephen J.
; TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL
; FILE OF INVENTION: DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND
; FILE REFERENCE: USING SAME
; CURRENT APPLICATION NUMBER: 227
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/630,915A
; FILING DATE: 03-APR-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Mirock, S. Leslie
; REGISTRATION NUMBER: 18,972
; REFERENCE/DOCKET NUMBER: 1101-174
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 111:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 55 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
US-08-630-915A-111

Query Match          10.3%; Score 43.5; DB 4; Length 55;
Best Local Similarity 41.4%; Pred. No. 94;
Matches 12; Conservative 8; Mismatches 4; Indels 5; Gaps 2;

QY 16 SISNSLVAMDFS-GQKSRVIENTEALS 43
Db 23 TVNKGSLVALGFGDQGEAR----PEELN 47

RESULT 11
US-09-267-177-12
; Sequence 12, Application US/09267177
; Patent No. 6287856
; GENERAL INFORMATION:
; APPLICANT: Poet, Steven E.
; APPLICANT: Ritchie, Branson W.
; APPLICANT: Niagro, Frank D.
; APPLICANT: Lukert, Phil D.
; TITLE OF INVENTION: Vaccines against Circovirus Infections
; FILE REFERENCE: 21099,0057
; CURRENT APPLICATION NUMBER: US/09/267,177
; CURRENT FILING DATE: 1999-03-12
; EARLIER APPLICATION NUMBER: 60/077,890
; EARLIER FILING DATE: 1998-03-13
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 12
; LENGTH: 74

```

RESULT 13  
US-09-247-155-173  
; Sequence 173, Application US/09247155A  
; Patent No. 6312922

REFERENCE/DOCKET NUMBER: 1101-174  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-8864/9741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 208:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 61 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-08-630-915A-208

Query Match 10.0%; Score 42.5; DB 4; Length 61;  
Best Local Similarity 33.3%; Pred. No. 1.5e-02;  
Matches 10; Conservative 6; Mismatches 13; Indels 1; Gaps 1;

QY 9 QSPMRISSENSLVAMDFSGKSRVIENP 38  
DB 6 QTLXPFSVTEELNEFE-KGETMEVIEKP 34

RESULT 15  
US-07-641-971B-5  
; Sequence 5, Application US/07641971B  
; Patent No. 5236706  
; GENERAL INFORMATION:  
; APPLICANT: Debre, Patrice  
; APPLICANT: Mossalayi, Mohammed D  
; TITLE OF INVENTION: A PHARMACEUTICAL PREPARATION FOR THE  
; TITLE OF INVENTION: MATURATION OF PROTHYMOCYTES  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Irving M. Fishman, CIBA-GEIGY Corporation  
; STREET: 556 Morris Avenue  
; CITY: Summit  
; STATE: New Jersey  
; COUNTRY: USA  
; ZIP: 07901  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/641.971B  
; FILING DATE: 19910116  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 90016254  
; FILING DATE: 24-JAN-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Fishman, Irving M  
; REGISTRATION NUMBER: 30258  
; REFERENCE/DOCKET NUMBER: 4-17921/+/DEB  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 908-277-4832  
; TELEFAX: 908-277-4306  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 40 amino acids  
; TYPE: AMINO ACID  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; FRAGMENT TYPE: N-terminal  
US-07-641-971B-5

Query Match 9.9%; Score 42; DB 1; Length 40;  
Best Local Similarity 37.9%; Pred. No. 94;  
Matches 11; Conservative 6; Mismatches 4; Indels 1; Gaps 1;

Matches 11; Conservative 6; Mismatches 8; Indels 4; Gaps 1;

QY 12 SPMRISSENSLVAMDFSGKSRVIENPTE 40  
DB 1 APVRSLN-----CTLRDGGKSLVMSGPYE 25

RESULT 16  
US-07-781-248A-5  
; Sequence 5, Application US/07781248A  
; Patent No. 5246699  
; GENERAL INFORMATION:  
; APPLICANT: Debre, Patrice  
; APPLICANT: Mossalayi, Mohammed D  
; TITLE OF INVENTION: MATURATION OF HEMATOPOIETIC CELLS  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Irving M. Fishman, CIBA-GEIGY Corporation  
; STREET: 556 Morris Avenue  
; CITY: Summit  
; STATE: New Jersey  
; COUNTRY: USA  
; ZIP: 07901  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/781.248A  
; FILING DATE: 19911230  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 90103565  
; FILING DATE: 09-MAY-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ikeler, Barbara J.  
; REGISTRATION NUMBER: 36,170  
; REFERENCE/DOCKET NUMBER: 4-18065/A/DEB  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 908-277-3368  
; TELEFAX: 908-277-4306  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 40 amino acids  
; TYPE: AMINO ACID  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; FRAGMENT TYPE: N-terminal  
US-07-781-248A-5

Query Match 9.9%; Score 42; DB 1; Length 40;  
Best Local Similarity 37.9%; Pred. No. 94;  
Matches 11; Conservative 6; Mismatches 8; Indels 4; Gaps 1;

QY 12 SPMRISSENSLVAMDFSGKSRVIENPTE 40  
DB 1 APVRSLN-----CTLRDGGKSLVMSGPYE 25

RESULT 17  
US-08-776-059-16  
; Sequence 16, Application US/08776059B  
; Patent No. 6271368  
; GENERAL INFORMATION:  
; APPLICANT: LENTZEN, Hans  
; APPLICANT: ECK, Jurgen  
; APPLICANT: BAUR, Axel  
; APPLICANT: ZINKE, Holger  
; TITLE OF INVENTION: RECOMBINANT MISTLETOE LECTIN (RML)



```

; LENGTH: 84
5171684-3
Query Match          9%; Score 42: DB 6: Length 84;
Best Local Similarity 26.5%; Pred. No. 2.Be+02;
Matches 13; Conservative 19; Indels
QOY      6 C$SOISIP-----MSISENSILVAMDSGKSRVINEPTEALSVAVEE 48
DBB      38 CVNRVRAPREGVMRKHSRTLFPRDWTIAESGL--NPTEVIDVVEE 84
|:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::|
|:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::| |:::|
RESULT 20
US-09-023-082A-76
Sequence 76 Application US/09023082A
Patent No. 6077692
GENERAL INFORMATION:
APPLICANT: RUBEN STEVEN M.
APPLICANT: JHENEZ, PABLO
APPLICANT: DUAN, D. ROXANNE
APPLICANT: RAMPI, MARK A.
APPLICANT: MENDRICK, DONNA
APPLICANT: ZHANG, JUN
APPLICANT: NI, JIAN
APPLICANT: MOORE, PAUL A.
APPLICANT: COLEMAN, TIMOTHY A.
APPLICANT: GRUBER, JOACHIM R.
APPLICANT: DILLON, PATRICK J.
APPLICANT: GENTZ, REINER L.
TITLE OF INVENTION: KERATINOCYTE GROWTH FACTOR-2
NUMBER OF SEQUENCES: 148
CORRESPONDENCE ADDRESS:
ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C.
STREET: 1100 NEW YORK AVE, NW, SUITE 600
CITY: WASHINGTON
STATE: DC
COUNTRY: USA
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/023,082A
FILING DATE: 13-FEB-1998
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/01790
FILING DATE: 14-FEB-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/461,195
FILING DATE: 05-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/023,852
FILING DATE: 13-AUG-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/039,045
FILING DATE: 28-FEB-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/862,432
FILING DATE: 23-MAY-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/910,875
FILING DATE: 13-AUG-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/055,561
FILING DATE: 13-AUG-1997
ATTORNEY/AGENT INFORMATION:
NAME: STEFFE, ERIC K.
REGISTRATION NUMBER: 36,688
REFERENCE/DOCKET NUMBER: 1488. 0360008/EKS
TELECOMMUNICATION INFORMATION:

```

```
/
/ TELEPHONE: 202-371-2600
/ TELEFAX: 202-371-2540
/ INFORMATION FOR SEQ ID NO: 76:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 87 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ US-09-023-082A-76

Query Match          9.8%; Score 41.5; DB 3; Length 87;
Best Local Similarity 19.3%; Pred. No. 3.4e+02;
Matches 16; Conservative 15; Mismatches 31; Indels 21; Gaps 3;

QY 14 MRSISESLVAMDFSGQ-----KSRVIENPTSEALSVAVEEGLAWRKKG---C 57
DB 2 VKAINSNYILAMNKKGLYGSREFNNDCKLKERIEENGNTYA-----SFNQHNGROMY 56

QY 58 LRLGTHGSPASSOSSATNMAIH 80
DB 57 VALNGKGAPRRGQKTRKNTSAH 79

RESULT 21
US-08-927-219-49
; Sequence 49, Application US/08927219
; Patent No. 6187533
; GENERAL INFORMATION:
; APPLICANT: Bell, Graeme I.
; APPLICANT: Yamagata, Kazuya
; APPLICANT: Oda, Naohisha
; APPLICANT: Kaisaki, Pamela J.
; APPLICANT: Furuta, Hiroto
; APPLICANT: Horikawa, Yukio
; APPLICANT: Menzel, Stephen
; TITLE OF INVENTION: MUTATIONS IN THE DIABETES SUSCEPTIBILITY
; TITLE OF INVENTION: GENES HEPATOCYTE NUCLEAR FACTOR (HNF) 1 ALPHA, HNF-1BETA
; TITLE OF INVENTION: AND HNF-4ALPHA
; NUMBER OF SEQUENCES: 147
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P. O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/927,219
; FILING DATE: Concurrently Herewith
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/029,679
; FILING DATE: 30-OCT-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/028,056
; FILING DATE: 02-OCT-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/025,719
; FILING DATE: 10-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Wilson, Mark B.
; REGISTRATION NUMBER: 37,259
; REFERENCE/DOCKET NUMBER: ARCD.272
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 512/418-3000
; TELEFAX: 512/474-7577
; INFORMATION FOR SEQ ID NO: 49:
; SEQUENCE CHARACTERISTICS:
/
/ LENGTH: 51 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ US-08-927-219-49

Query Match          9.7%; Score 41; DB 4; Length 51;
Best Local Similarity 28.2%; Pred. No. 1.8e+02;
Matches 11; Conservative 3; Mismatches 11; Indels 14; Gaps 1;

QY 51 AWRKKG-----LRLGTHGSPASSOSSAT 75
DB 12 SWRRAGCPYASLTWCCPSRCSRMTMTSKPSSSLT 50

RESULT 22
US-09-246-500B-9
; Sequence 9, Application US/09246500B
; Patent No. 6235494
; GENERAL INFORMATION:
; APPLICANT: Hugli, Tony E.
; TITLE OF INVENTION: Substrates for Assessing Mannan-Binding
; TITLE OF INVENTION: Protein-Associated Serine Protease Activity and Methods
; TITLE OF INVENTION: Using the Substrates
; FILE REFERENCE: 24730-2204
; CURRENT APPLICATION NUMBER: US/09/246,500B
; CURRENT FILING DATE: 1999-02-08
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 76
; TYPE: PRT
; ORGANISM: Mouse C4a Anaphylatoxin
; FEATURE:
; NAME/KEY: CARBOHYD
; LOCATION: (66)...(68)
; US-09-246-500B-9

Query Match          9.7%; Score 41; DB 4; Length 76;
Best Local Similarity 38.5%; Pred. No. 3.3e+02;
Matches 10; Conservative 4; Mismatches 10; Indels 2; Gaps 1;

QY 59 RLGTHGSPAS--SOSATNMAIHS 82
DB 11 KLGQYASPTAKRCQDGVTRLPMKRS 36

RESULT 23
US-08-459-568-52
; Sequence 52, Application US/08459568
; Patent No. 5811304
; GENERAL INFORMATION:
; APPLICANT: Huang, Shi
; TITLE OF INVENTION: Retinoblastoma Protein - Interacting
; TITLE OF INVENTION: Zinc Finger Proteins
; NUMBER OF SEQUENCES: 93
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/459,568
; FILING DATE: 02-JUN-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
```

```

; APPLICATION NUMBER: US 08/399,411
; FILING DATE: 06-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 1264
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 52:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 66 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
;
US-08-459-568-52

```

```

Query Match          9.6%; Score 40.5; DB 2; Length 66;
Best Local Similarity 47.4%; Pred. No. 3.1e+02;
Matches 9; Conservative 7; Mismatches 2; Indels 1; Gaps 1;

```

```

QY 16 SISSENSLVAMDFS-GOKSR 33
Db 22 TVNKGSLVALGFSGDGEAR 40

```

```

RESULT 24
US-08-399-411-52
; Sequence 52, Application US/08399411
; Patent No. 5831008
; GENERAL INFORMATION:
; APPLICANT: Huang, Shi
; TITLE OF INVENTION: Retinoblastoma Protein - Interacting
; NUMBER OF SEQUENCES: 93
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/399,411
; FILING DATE: 06-MAR-1995
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 1264
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 52:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 66 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
;
US-08-399-411-52

```

```

Query Match          9.6%; Score 40.5; DB 2; Length 66;
Best Local Similarity 47.4%; Pred. No. 3.1e+02;
Matches 9; Conservative 7; Mismatches 2; Indels 1; Gaps 1;

```

```

QY 16 SISSENSLVAMDFS-GOKSR 33
Db 22 TVNKGSLVALGFSGDGEAR 40

```

```

RESULT 25
US-08-516-859A-52
; Sequence 52, Application US/08516859A
; Patent No. 6069231
; GENERAL INFORMATION:
; APPLICANT: Huang, Shi
; TITLE OF INVENTION: Retinoblastoma Protein - Interacting
; NUMBER OF SEQUENCES: 106
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/516,859A
; FILING DATE: 18-AUG-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/399,411
; FILING DATE: 06-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/292,683
; FILING DATE: 18-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 1776
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 52:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 66 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
;
US-08-516-859A-52

```

```

Query Match          9.6%; Score 40.5; DB 3; Length 66;
Best Local Similarity 47.4%; Pred. No. 3.1e+02;
Matches 9; Conservative 7; Mismatches 2; Indels 1; Gaps 1;

```

```

QY 16 SISSENSLVAMDFS-GOKSR 33
Db 22 TVNKGSLVALGFSGDGEAR 40

```

```

Search completed: March 28, 2003, 09:09:39
Job time : 32.4724 secs

```

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 28, 2003, 08:55:25 ; Search time 66.1417 Seconds  
(without alignments)  
169.228 Million cell updates/sec

Sequence: 1 QCRSCSSQISPMRSISEN.....SPTASSQSSATNMAIHRSP 84

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 160864

Minimum DB seq length: 40  
Maximum DB seq length: 90

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 1000 summaries

No.	Score	Match	Length	ID	Description
1	386	91.3	84	21 AAB18942	Peptide derived fr
2	386	91.3	84	21 AAB18938	Peptide derived fr
3	212	50.1	43	21 AAB18941	Peptide derived fr
4	205	48.5	43	21 AAB18937	Peptide derived fr
5	191	45.2	80	21 AAB18954	Peptide derived fr
6	191	45.2	80	21 AAB18952	Peptide derived fr
7	189	44.7	82	21 AAB18950	Peptide derived fr
8	186	44.0	82	21 AAB18946	Peptide derived fr
9	179	42.3	80	21 AAB18958	Peptide derived fr
10	169	40.0	43	21 AAB18949	Peptide derived fr

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result	No.	Score	Query Match	Length	DB	ID	Description
1	423	100.0	84	21	AAB18942	Peptide derived fr	
2	386	91.3	84	21	AAB18938	Peptide derived fr	
3	212	50.1	43	21	AAB18941	Peptide derived fr	
4	205	48.5	43	21	AAB18937	Peptide derived fr	
5	191	45.2	80	21	AAB18954	Peptide derived fr	
6	191	45.2	80	21	AAB18952	Peptide derived fr	
7	189	44.7	82	21	AAB18950	Peptide derived fr	
8	186	44.0	82	21	AAB18946	Peptide derived fr	
9	179	42.3	80	21	AAB18958	Peptide derived fr	
10	169	40.0	43	21	AAB18949	Peptide derived fr	

11	162	38.3	43	21 AAB18957	Peptide derived fr
12	161	38.1	43	21 AAB18945	Peptide derived fr
13	159	37.6	43	21 AAB18953	Peptide derived fr
14	159	37.6	43	21 AAB18961	Peptide derived fr
15	65.5	15.5	47	22 AAO02215	Human polypeptide
16	61	14.4	77	22 AAO13559	Human polypeptide
17	59.5	14.1	76	22 AAO09950	Human polypeptide
18	58.5	13.8	78	22 AAO11222	Novel human diagno
19	58.5	13.8	78	22 ABO07920	Human polypeptide
20	56.5	13.4	48	22 AAO04576	Human polypeptide
21	56	13.2	72	22 AAO13082	Novel human secret
22	56	13.2	72	22 ABO05212	Human albumin fusi
23	56	13.2	79	22 ABO091124	Human secreted pro
24	55.5	13.1	49	22 AAO00883	Human polypeptide
25	55	13.0	73	22 AAO06913	Human polypeptide
26	54.5	12.9	53	22 ABO11592	Human PRO0663 homo
27	54.5	12.9	57	22 ABO02340	Human polypeptide
28	54.5	12.9	60	21 AAB53484	Human colon cancer
29	54.5	12.9	72	21 AAG00338	Human polypeptide
30	54	12.8	51	22 AAO10414	Human polypeptide
31	54	12.8	86	22 AAG76449	Human colon cancer
32	53	12.5	49	22 AAO13029	Human polypeptide
33	53	12.5	70	22 AAO00993	Human polypeptide
34	53	12.5	75	22 AAU49021	Propionibacterium
35	52.5	12.4	52	22 AAO05395	Human polypeptide
36	52.5	12.4	67	22 AAO10683	Human polypeptide
37	52.5	12.4	75	22 AAO09630	Human polypeptide
38	52	12.3	49	22 AAO05487	Human polypeptide
39	52	12.3	53	22 AAU20221	Human novel endocr
40	52	12.3	70	21 ABB17410	Human secreted pro
41	52	12.3	78	22 ABB17410	Human nervous syst
42	52	12.3	90	22 AAO08661	Novel human secret
43	51.5	12.2	61	22 AAO08661	Human polypeptide
44	51.5	12.2	64	22 AAO08661	Human polypeptide
45	51.5	12.2	68	22 AAO08639	Human immune/haema
46	51.5	12.2	72	22 AAG03340	Human secreted pro
47	51.5	12.2	74	22 AAB64880	Human polypeptide
48	51.5	12.2	83	22 ABO06480	Human polypeptide
49	51.5	12.2	83	22 ABO06480	Human reproductive
50	51.5	12.2	83	22 ABO06480	Human ovarian anti
51	51.5	12.2	83	23 ABO06480	Novel ovarian rela
52	51.5	12.2	88	23 ABO01560	Human ORF3 protein
53	51	12.1	57	22 AAO08958	Human polypeptide
54	51	12.1	62	22 AAU29728	Novel human secret
55	51	12.1	72	22 AAO05954	Human reproductive
56	51	12.1	83	22 AAO05954	Human polypeptide
57	51	12.1	84	23 ABO05954	Human polypeptide
58	51	12.1	87	22 ABO05954	Human peptide enco
59	51	12.1	88	22 ABO05954	Human polypeptide
60	50.5	11.9	55	21 AAG03473	Human secreted pro
61	50.5	11.9	69	22 AAO08892	Human polypeptide
62	50.5	11.9	71	22 AAU43364	Propionibacterium
63	50.5	11.9	76	22 AAO05936	Propionibacterium
64	50.5	11.9	81	22 AAO02479	Human polypeptide
65	50.5	11.9	83	23 ABO02479	Human polypeptide
66	50.5	11.9	88	23 ABO02479	Human ovarian anti
67	50	11.8	48	22 AAO09332	Human ORF35 protei
68	50	11.8	53	22 AAO09332	Human polypeptide
69	50	11.8	60	22 AAO09332	Human immune/haema
70	50	11.8	74	22 AAO09332	Novel human secret
71	50	11.8	77	22 AAO09332	Human polypeptide
72	50	11.8	78	22 AAO09332	Human immune/haema
73	49.5	11.7	45	21 AAB34648	Propionibacterium
74	49.5	11.7	48	22 AAB34648	Human secreted pro
75	49.5	11.7	49	22 ABO05087	Human polypeptide
76	49.5	11.7	50	22 ABO05087	Human nervous syst
77	49.5	11.7	52	22 ABO05087	Human polypeptide
78	49.5	11.7	56	22 AAO06903	Propionibacterium
79	49.5	11.7	57	22 AAO11767	Human polypeptide
80	49.5	11.7	74	21 AAG00397	Human secreted pro
81	49.5	11.7	80	23 ABO0397	Human albumin fusi
82	49.5	11.7	80	23 ABO0397	Human polypeptide
83	49.5	11.7	80	23 ABO0397	Human gene 9 encod
84	49.5	11.7	80	23 ABO0397	Human gene 9 encod















960 39.5 9.3 73 21 AAB36290  
 961 39.5 9.3 73 22 AAU52874  
 962 39.5 9.3 73 22 AAU52874  
 963 39.5 9.3 73 22 AAU52874  
 964 39.5 9.3 73 22 AAU52874  
 965 39.5 9.3 73 22 AAU52874  
 966 39.5 9.3 73 22 AAU52874  
 967 39.5 9.3 73 22 AAU52874  
 968 39.5 9.3 73 22 AAU52874  
 969 39.5 9.3 73 22 AAU52874  
 970 39.5 9.3 73 22 AAU52874  
 971 39.5 9.3 73 22 AAU52874  
 972 39.5 9.3 73 22 AAU52874  
 973 39.5 9.3 73 22 AAU52874  
 974 39.5 9.3 73 22 AAU52874  
 975 39.5 9.3 73 22 AAU52874  
 976 39.5 9.3 73 22 AAU52874  
 977 39.5 9.3 73 22 AAU52874  
 978 39.5 9.3 73 22 AAU52874  
 979 39.5 9.3 73 22 AAU52874  
 980 39.5 9.3 73 22 AAU52874  
 981 39.5 9.3 73 22 AAU52874  
 982 39.5 9.3 73 22 AAU52874  
 983 39.5 9.3 73 22 AAU52874  
 984 39.5 9.3 73 22 AAU52874  
 985 39.5 9.3 73 22 AAU52874  
 986 39.5 9.3 73 22 AAU52874  
 987 39.5 9.3 73 22 AAU52874  
 988 39.5 9.3 73 22 AAU52874  
 989 39.5 9.3 73 22 AAU52874  
 990 39.5 9.3 73 22 AAU52874  
 991 39.5 9.3 73 22 AAU52874  
 992 39.5 9.3 73 22 AAU52874  
 993 39.5 9.3 73 22 AAU52874  
 994 39.5 9.3 73 22 AAU52874  
 995 39.5 9.3 73 22 AAU52874  
 996 39.5 9.3 73 22 AAU52874  
 997 39.5 9.3 73 22 AAU52874  
 998 39.5 9.3 73 22 AAU52874  
 999 39.5 9.3 73 22 AAU52874  
 1000 39.5 9.3 73 22 AAU52874

XX WPI: 2000-587566/55.  
 XX Fragments of Grb family proteins to identify compounds are useful in  
 XX treating insulin-associated diseases, particularly diabetes and obesity  
 XX  
 XX Claim 2; Page 26; 46pp; French.  
 XX  
 XX B18937-64 represent the PIR (phosphorylated insulin receptor interacting  
 XX region) or PIR-SH2 (Src homology 2) domains of a Grb7 family protein.  
 XX PIR is the actual binding region but its effect is about 10 times  
 XX greater in presence of SH2 (which by itself is inactive). Agents that  
 XX affect binding between the peptides and the insulin receptor can  
 XX stimulate or inhibit tyrosine kinase activity of the receptor. The  
 XX peptides are used for screening molecules for ability to treat diseases  
 XX in which insulin is implicated. The peptides are used to identify agents  
 XX that are potentially useful for treating insulin-associated diseases,  
 XX particularly diabetes and obesity but also polycystic ovarian syndrome  
 XX and syndrome X.  
 XX Sequence 84 AA;  
 XX

Query Match 100.0%; Score 423; DB 21; Length 84;  
 Best Local Similarity 100.0%; Pred. No. 6.2e-47; Indels 0; Caps 0;  
 Matches 84; Conservative 0; Mismatches 0;  
 QY 1 QGRSGCSQSISPMRSISNSLVAMDFSGQSRVIENTALSVAVEGLAWRRKGCLRL 60  
 DB 1 QGRSGCSQSISPMRSISNSLVAMDFSGQSRVIENTALSVAVEGLAWRRKGCLRL 60  
 QY 61 GTGSGPTASSQSATNMAIHRSP 84  
 DB 61 GTGSGPTASSQSATNMAIHRSP 84

RESULT 2  
 AAB18938  
 ID AAB18938 standard; peptide; 84 AA.  
 AC AAB18938;  
 AC AAB18938;  
 DT 08-FEB-2001 (first entry)

# ALIGNMENTS

RESULT 1  
 AAB18942  
 ID AAB18942 standard; peptide; 84 AA.  
 AC AAB18942;  
 DT 08-FEB-2001 (first entry)  
 XX Peptide derived from the PIR or PIR-SH2 domain of Grb7 protein.  
 XX  
 XX Phosphorylated insulin receptor interacting region; Grb7 family protein;  
 XX insulin receptor; tyrosine kinase; insulin; insulin-associated disease;  
 XX diabetes; obesity; polycystic ovarian syndrome; syndrome X.  
 XX  
 XX ( Homo sapiens  
 XX 2000-587566/55  
 XX 21-SEP-2000.  
 XX 14-MAR-2000; 2000MO-FR00613.  
 XX 15-MAR-1999; 99FR-0003159.  
 XX (CNRS ) CNRS CENT NAT RECH SCI.  
 XX Burnol A, Perdureau D, Kasus-Jacobi A, Bereziat V, Girard J;  
 XX WPI: 2000-587566/55.  
 XX Fragments of Grb family proteins to identify compounds are useful in  
 XX treating insulin-associated diseases, particularly diabetes and obesity  
 XX  
 XX Claim 2; Page 23-24; 46pp; French.  
 XX B18937-64 represent the PIR (phosphorylated insulin receptor interacting

CC region) or PIR-SH2 (Src homology 2) domains of a Grb7 family protein.  
 CC PIR is the actual binding region but its effect is about 10 times  
 CC greater in presence of SH2 (which by itself is inactive). Agents that  
 CC affect binding between the peptides and the insulin receptor can  
 CC stimulate or inhibit tyrosine kinase activity of the receptor. The  
 CC peptides are used for screening molecules for ability to treat diseases  
 CC in which insulin is implicated. The peptides are used to identify agents  
 CC that are potentially useful for treating insulin-associated diseases,  
 CC particularly diabetes and obesity but also polycystic ovarian syndrome  
 CC and syndrome X.  
 CC  
 XX Sequence 84 AA;  
 SQ  
 Query Match 91.3%; Score 386; DB 21; Length 84;  
 Best Local Similarity 88.1%; Pred. No. 3.9e-42;  
 Matches 74; Conservative 5; Mismatches 5; Indels 0; Gaps 0;  
 QY 1 QGRSGCSQSISPMRSISNSLVAMDFSGQSRVIENTEALSVAVEEGLAWRKGCLRL 60  
 DB 1 QARSACSQSVSPMRVSNSLVAMDFSGQKTRVIDNPTLEALSVAVEEGLAWRKGCLRL 60  
 QY 61 GTHGSPTASSQSASATNMAIHRSQP 84  
 DB 61 GNHGSPTAPSSQSSAVNMAIHRSQP 84  
 RESULT 3  
 AAB18941  
 ID AAB18941 standard; peptide; 43 AA.  
 XX  
 AC AAB18941;  
 XX  
 DT 08-FEB-2001 (first entry)  
 XX  
 DE Peptide derived from the PIR or PIR-SH2 domain of Grb7 protein.  
 XX  
 KW Phosphorylated insulin receptor interacting region; Grb7 family protein;  
 KW insulin receptor; tyrosine kinase; insulin; insulin-associated disease;  
 KW diabetes; obesity; polycystic ovarian syndrome; syndrome X.  
 OS Homo sapiens.  
 XX  
 PN WO200055634-A1.  
 XX  
 PD 21-SEP-2000.  
 XX  
 PF 14-MAR-2000; 2000WO-FR00613.  
 XX  
 PR 15-MAR-1999; 99FR-0003159.  
 XX  
 PA (CNRS ) CNRS CENT NAT RECH SCI.  
 XX  
 PI Burnol A, Perdureau D, Kasus-Jacobi A, Bereziat V, Girard J;  
 XX  
 DR WPI; 2000-587566/55.  
 XX  
 PT Fragments of Grb family proteins to identify compounds are useful in  
 PT treating insulin-associated diseases, particularly diabetes and obesity  
 XX  
 PS Claim 2; Page 25; 46pp; French.  
 XX  
 CC Bl937-64 represent the PIR (phosphorylated insulin receptor interacting  
 CC region) or PIR-SH2 (Src homology 2) domains of a Grb7 family protein.  
 CC PIR is the actual binding region but its effect is about 10 times  
 CC greater in presence of SH2 (which by itself is inactive). Agents that  
 CC affect binding between the peptides and the insulin receptor can  
 CC stimulate or inhibit tyrosine kinase activity of the receptor. The  
 CC peptides are used for screening molecules for ability to treat diseases  
 CC in which insulin is implicated. The peptides are used to identify agents  
 CC that are potentially useful for treating insulin-associated diseases,  
 CC particularly diabetes and obesity but also polycystic ovarian syndrome  
 CC and syndrome X.

XX  
 SQ Sequence 43 AA;  
 Query Match 50.1%; Score 212; DB 21; Length 43;  
 Best Local Similarity 100.0%; Pred. No. 5.3e-20;  
 Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 13 PMRSISNSLVAMDFSGQSRVIENTEALSVAVEEGLAWRKG 55  
 DB 1 PMRSISNSLVAMDFSGQSRVIENTEALSVAVEEGLAWRKG 43  
 RESULT 4  
 AAB18937  
 ID AAB18937 standard; peptide; 43 AA.  
 XX  
 AC AAB18937;  
 XX  
 DT 08-FEB-2001 (first entry)  
 XX  
 DE Peptide derived from the PIR or PIR-SH2 domain of Grb7 protein.  
 XX  
 KW Phosphorylated insulin receptor interacting region; Grb7 family protein;  
 KW insulin receptor; tyrosine kinase; insulin; insulin-associated disease;  
 KW diabetes; obesity; polycystic ovarian syndrome; syndrome X.  
 OS Rattus sp.  
 XX  
 PN WO200055634-A1.  
 XX  
 PD 21-SEP-2000.  
 XX  
 PF 14-MAR-2000; 2000WO-FR00613.  
 XX  
 PR 15-MAR-1999; 99FR-0003159.  
 XX  
 PA (CNRS ) CNRS CENT NAT RECH SCI.  
 XX  
 PI Burnol A, Perdureau D, Kasus-Jacobi A, Bereziat V, Girard J;  
 XX  
 DR WPI; 2000-587566/55.  
 XX  
 PT Fragments of Grb family proteins to identify compounds are useful in  
 PT treating insulin-associated diseases, particularly diabetes and obesity  
 XX  
 PS Claim 2; Page 23; 46pp; French.  
 XX  
 CC Bl937-64 represent the PIR (phosphorylated insulin receptor interacting  
 CC region) or PIR-SH2 (Src homology 2) domains of a Grb7 family protein.  
 CC PIR is the actual binding region but its effect is about 10 times  
 CC greater in presence of SH2 (which by itself is inactive). Agents that  
 CC affect binding between the peptides and the insulin receptor can  
 CC stimulate or inhibit tyrosine kinase activity of the receptor. The  
 CC peptides are used for screening molecules for ability to treat diseases  
 CC in which insulin is implicated. The peptides are used to identify agents  
 CC that are potentially useful for treating insulin-associated diseases,  
 CC particularly diabetes and obesity but also polycystic ovarian syndrome  
 CC and syndrome X.  
 XX  
 SQ Sequence 43 AA;  
 Query Match 48.5%; Score 205; DB 21; Length 43;  
 Best Local Similarity 93.0%; Pred. No. 4.3e-19;  
 Matches 40; Conservative 3; Mismatches 0; Indels 0; Gaps 0;  
 QY 13 PMRSISNSLVAMDFSGQSRVIENTEALSVAVEEGLAWRKG 55  
 DB 1 PMRSISNSLVAMDFSGQKTRVIDNPTLEALSVAVEEGLAWRKG 43  
 RESULT 5  
 AAB18954

ID AAB18954 standard; peptide; 80 AA.  
 AC AAB18954;  
 DT 08-FEB-2001 (first entry)  
 DE Peptide derived from the PIR or PIR-SH2 domain of Grb7 protein.  
 KW Phosphorylated insulin receptor interacting region; Grb7 family protein;  
 KW insulin receptor; tyrosine kinase; insulin; insulin-associated disease;  
 KW diabetes; obesity; polycystic ovarian syndrome; syndrome X.  
 OS Rattus sp.  
 PN WO200055634-A1.  
 PD 21-SEP-2000.  
 XX 14-MAR-2000; 2000WO-FR00613.  
 XX 15-MAR-1999; 99FR-0003159.  
 XX (CNRS ) CNRS CENT NAT RECH SCI.  
 XX Burnol A, Perdereau D, Kasus-Jacobi A, Bereziat V, Girard J;  
 XX WPI; 2000-587566/55.  
 XX Fragments of Grb family proteins to identify compounds are useful in  
 XX treating insulin-associated diseases, particularly diabetes and obesity  
 XX Claim 2; Page 32; 46pp; French.  
 XX B18937-64 represent the PIR (phosphorylated insulin receptor interacting  
 XX region) or PIR-SH2 (Src homology 2) domains of a Grb7 family protein.  
 XX PIR is the actual binding region but its effect is about 10 times  
 XX greater in presence of SH2 (which by itself is inactive). Agents that  
 XX affect binding between the peptides and the insulin receptor can  
 XX stimulate or inhibit tyrosine kinase activity of the receptor. The  
 XX peptides are used for screening molecules for ability to treat diseases  
 XX in which insulin is implicated. The peptides are used to identify agents  
 XX that are potentially useful for treating insulin-associated diseases,  
 XX particularly diabetes and obesity but also polycystic ovarian syndrome  
 XX and syndrome X.  
 XX Sequence 80 AA;  
 SQ Query Match 45.2%; Score 191; DB 21; Length 80;  
 Best Local Similarity 59.7%; Pred. No. 6.9e-17;  
 Matches 43; Conservative 8; Mismatches 17; Indels 4; Gaps 2;  
 QY 13 PMRSISENSLVAMDFSGKSRVIENPTEALSVAVEGLAWRKKGCLRLGTHGSPASSQS 72  
 DB 13 PLRSVSDNTLVAMDFSGHAGRVIDNPREALSAAMEEAQAWRKKNHRLSL---PTTCSGS 69  
 QY 73 SATNMAIHRSP 84  
 DB 70 S-LSAAIHRTP 80  
 RESULT 6  
 ID AAB18962 standard; peptide; 80 AA.  
 AC AAB18962;  
 DT 08-FEB-2001 (first entry)  
 DE Peptide derived from the PIR or PIR-SH2 domain of Grb7 protein.  
 KW Phosphorylated insulin receptor interacting region; Grb7 family protein;  
 KW insulin receptor; tyrosine kinase; insulin; insulin-associated disease;  
 KW diabetes; obesity; polycystic ovarian syndrome; syndrome X.

KW diabetes; obesity; polycystic ovarian syndrome; syndrome X.  
 XX Mus muris.  
 OS WO200055634-A1.  
 PN 21-SEP-2000.  
 XX 14-MAR-2000; 2000WO-FR00613.  
 XX 15-MAR-1999; 99FR-0003159.  
 XX (CNRS ) CNRS CENT NAT RECH SCI.  
 XX Burnol A, Perdereau D, Kasus-Jacobi A, Bereziat V, Girard J;  
 XX WPI; 2000-587566/55.  
 XX Fragments of Grb family proteins to identify compounds are useful in  
 XX treating insulin-associated diseases, particularly diabetes and obesity  
 XX Claim 2; Page 37; 46pp; French.  
 XX B18937-64 represent the PIR (phosphorylated insulin receptor interacting  
 XX region) or PIR-SH2 (Src homology 2) domains of a Grb7 family protein.  
 XX PIR is the actual binding region but its effect is about 10 times  
 XX greater in presence of SH2 (which by itself is inactive). Agents that  
 XX affect binding between the peptides and the insulin receptor can  
 XX stimulate or inhibit tyrosine kinase activity of the receptor. The  
 XX peptides are used for screening molecules for ability to treat diseases  
 XX in which insulin is implicated. The peptides are used to identify agents  
 XX that are potentially useful for treating insulin-associated diseases,  
 XX particularly diabetes and obesity but also polycystic ovarian syndrome  
 XX and syndrome X.  
 XX Sequence 80 AA;  
 SQ Query Match 45.2%; Score 191; DB 21; Length 80;  
 Best Local Similarity 59.7%; Pred. No. 6.9e-17;  
 Matches 43; Conservative 8; Mismatches 17; Indels 4; Gaps 2;  
 QY 13 PMRSISENSLVAMDFSGKSRVIENPTEALSVAVEGLAWRKKGCLRLGTHGSPASSQS 72  
 DB 13 PLRSVSDNTLVAMDFSGHAGRVIDNPREALSAAMEEAQAWRKKNHRLSL---PTTCSGS 69  
 QY 73 SATNMAIHRSP 84  
 DB 70 S-LSAAIHRTP 80  
 RESULT 7  
 ID AAB18950 standard; peptide; 82 AA.  
 AC AAB18950;  
 DT 08-FEB-2001 (first entry)  
 DE Peptide derived from the PIR or PIR-SH2 domain of Grb7 protein.  
 KW Phosphorylated insulin receptor interacting region; Grb7 family protein;  
 KW insulin receptor; tyrosine kinase; insulin; insulin-associated disease;  
 KW diabetes; obesity; polycystic ovarian syndrome; syndrome X.  
 OS Homo sapiens.  
 PN WO200055634-A1.  
 PD 21-SEP-2000.  
 XX 14-MAR-2000; 2000WO-FR00613.

PT 15-MAR-1999; 99FR-0003159.  
 XX (CNRS ) CNRS CENT NAT RECH SCI.  
 PA Burnol A, Perdureau D, Kasus-Jacobi A, Bereziat V, Girard J;  
 PI WPI; 2000-587566/55.  
 XX Fragments of Grb family proteins to identify compounds are useful in  
 PT treating insulin-associated diseases, particularly diabetes and obesity  
 XX  
 XX Claim 2; Page 30; 46pp; French.  
 XX BI8937-64 represent the PIR (phosphorylated insulin receptor interacting  
 CC region) or PIR-SH2 (Src homology 2) domains of a Grb7 family protein.  
 CC PIR is the actual binding region but its effect is about 10 times  
 CC greater in presence of SH2 (which by itself is inactive). Agents that  
 CC affect binding between the peptides and the insulin receptor can  
 CC stimulate or inhibit tyrosine kinase activity of the receptor. The  
 CC peptides are used for screening molecules for ability to treat diseases  
 CC in which insulin is implicated. The peptides are used to identify agents  
 CC that are potentially useful for treating insulin-associated diseases,  
 CC particularly diabetes and obesity but also polycystic ovarian syndrome  
 CC and syndrome X.  
 XX Sequence 82 AA;  
 XX Query Match 44.7%; Score 189; DB 21; Length 82;  
 XX Best Local Similarity 53.0%; Pred. No. 1.3e-16;  
 XX Matches 44; Conservative 11; Mismatches 26; Indels 2; Gaps 2;  
 OY 1 QGRSGCSQSISPMRSISNSLVAMDFSGQKSRVIENTALSVAVEGLAWRKGCLRL 60  
 DB 1 QQRKALLSPSTFVRSVNSLVAMDFSGQIGRVINPAQSALEEGHAWRRRS-TRM 59  
 OY 61 GTHGSPASSQSATNMAIHRSQ 83  
 DB 60 NILGQSPLHPSTLSTV-IHRTQ 81  
 RESULT 8  
 AAB18946  
 ID AAB18946 standard; peptide; 82 AA.  
 AC AAB18946;  
 DT 08-FEB-2001 (first entry)  
 XX Peptide derived from the PIR or PIR-SH2 domain of Grb7 protein.  
 XX Phosphorylated insulin receptor interacting region; Grb7 family protein;  
 KW insulin receptor; tyrosine kinase; insulin; insulin-associated disease;  
 KW diabetes; obesity; polycystic ovarian syndrome; syndrome X.  
 OS Mus muris.  
 XX WO200055634-A1.  
 PN 21-SEP-2000.  
 XX 14-MAR-2000; 2000WO-FR00613.  
 XX 15-MAR-1999; 99FR-0003159.  
 XX (CNRS ) CNRS CENT NAT RECH SCI.  
 XX Burnol A, Perdureau D, Kasus-Jacobi A, Bereziat V, Girard J;  
 PI WPI; 2000-587566/55.  
 XX Fragments of Grb family proteins to identify compounds are useful in  
 PT treating insulin-associated diseases, particularly diabetes and obesity  
 XX

PT Claim 2; Page 28; 46pp; French.  
 XX BI8937-64 represent the PIR (phosphorylated insulin receptor interacting  
 CC region) or PIR-SH2 (Src homology 2) domains of a Grb7 family protein.  
 CC PIR is the actual binding region but its effect is about 10 times  
 CC greater in presence of SH2 (which by itself is inactive). Agents that  
 CC affect binding between the peptides and the insulin receptor can  
 CC stimulate or inhibit tyrosine kinase activity of the receptor. The  
 CC peptides are used for screening molecules for ability to treat diseases  
 CC in which insulin is implicated. The peptides are used to identify agents  
 CC that are potentially useful for treating insulin-associated diseases,  
 CC particularly diabetes and obesity but also polycystic ovarian syndrome  
 CC and syndrome X.  
 XX Sequence 82 AA;  
 XX Query Match 44.0%; Score 186; DB 21; Length 82;  
 XX Best Local Similarity 54.1%; Pred. No. 3.2e-16;  
 XX Matches 46; Conservative 6; Mismatches 23; Indels 10; Gaps 3;  
 OY 3 RGCSSQSISPMRSISNSLVAMDFSGQKSRVIENTALSVAVEGLAWRKGCLRLGT 62  
 DB 3 RKGLPPFPNAPRSVNSLVAMDFSGQIGRVINPAQSALEEGHAWR-NGSTRMN- 60  
 OY 63 HGSPASSQS-----SATNMAIHRSQ 83  
 DB 61 -----ILSSQSLHPSTLNAVIRHTQ 81  
 RESULT 9  
 AAB18958  
 ID AAB18958 standard; peptide; 80 AA.  
 AC AAB18958;  
 DT 08-FEB-2001 (first entry)  
 XX Peptide derived from the PIR or PIR-SH2 domain of Grb7 protein.  
 XX Phosphorylated insulin receptor interacting region; Grb7 family protein;  
 KW insulin receptor; tyrosine kinase; insulin; insulin-associated disease;  
 KW diabetes; obesity; polycystic ovarian syndrome; syndrome X.  
 OS Homo sapiens.  
 XX WO200055634-A1.  
 PN 21-SEP-2000.  
 XX 14-MAR-2000; 2000WO-FR00613.  
 XX 15-MAR-1999; 99FR-0003159.  
 XX (CNRS ) CNRS CENT NAT RECH SCI.  
 XX Burnol A, Perdureau D, Kasus-Jacobi A, Bereziat V, Girard J;  
 PI WPI; 2000-587566/55.  
 XX Fragments of Grb family proteins to identify compounds are useful in  
 PT treating insulin-associated diseases, particularly diabetes and obesity  
 XX  
 XX Claim 2; Page 34-35; 46pp; French.  
 XX BI8937-64 represent the PIR (phosphorylated insulin receptor interacting  
 CC region) or PIR-SH2 (Src homology 2) domains of a Grb7 family protein.  
 CC PIR is the actual binding region but its effect is about 10 times  
 CC greater in presence of SH2 (which by itself is inactive). Agents that  
 CC affect binding between the peptides and the insulin receptor can  
 CC stimulate or inhibit tyrosine kinase activity of the receptor. The

CC peptides are used for screening molecules for ability to treat diseases  
 CC in which insulin is implicated. The peptides are used to identify agents  
 CC that are potentially useful for treating insulin-associated diseases,  
 CC particularly diabetes and obesity but also polycystic ovarian syndrome  
 CC and syndrome X.  
 XX  
 SQ

Sequence 80 AA;

Query Match 42.3%; Score 179; DB 21; Length 80;  
 Best Local Similarity 59.2%; Pred. No. 2.5e-15;  
 Matches 42; Conservative 8; Mismatches 17; Indels 4; Gaps 2;  
 OY 13 PMRSISENSLVAMDFSGKSRVIENPTEALSVAVEEGLAWRK 72  
 DB 13 PLRSASDNTLVAMDFSGHAGRVENPREALSVALEEAQAWRKKTNRHLSL--PMPASGT 69  
 OY 73 SATNMAIHRQ 83  
 DB 70 S-LSAAIHRQ 79

RESULT 10  
 AABI8949  
 ID AABI8949 standard; peptide: 43 AA.

XX AC AABI8949;  
 XX DT 08-FEB-2001 (first entry)  
 XX DE Peptide derived from the PIR or PIR-SH2 domain of Grb7 protein.  
 XX KW Phosphorylated insulin receptor interacting region; Grb7 family protein;  
 KW insulin receptor; tyrosine kinase; insulin; insulin-associated disease;  
 KW diabetes; obesity; polycystic ovarian syndrome; syndrome X.  
 XX OS Homo sapiens.  
 XX PN WO200055634-A1.  
 XX PD 21-SEP-2000.  
 XX PF 14-MAR-2000; 2000WO-FR00613.  
 XX PR 15-MAR-1999; 99FR-0003159.  
 XX PA (CNRS ) CNRS CENT NAT RECH SCI.  
 XX PI Burnol A, Perdureau D, Kasus-Jacobi A, Bereziat V, Girard J;  
 DR WPI; 2000-587566/55.  
 XX PT Fragments of Grb family proteins to identify compounds are useful in  
 PT treating insulin-associated diseases, particularly diabetes and obesity  
 PT -

Claim 2; Page 30; 46pp; French.

XX BI8937-64 represent the PIR (phosphorylated insulin receptor interacting  
 CC region) or PIR-SH2 (Src homology 2) domains of a Grb7 family protein.  
 CC PIR is the actual binding region but its effect is about 10 times  
 CC greater in presence of SH2 (which by itself is inactive). Agents that  
 CC affect binding between the peptides and the insulin receptor can  
 CC stimulate or inhibit tyrosine kinase activity of the receptor. The  
 CC peptides are used for screening molecules for ability to treat diseases  
 CC in which insulin is implicated. The peptides are used to identify agents  
 CC that are potentially useful for treating insulin-associated diseases,  
 CC particularly diabetes and obesity but also polycystic ovarian syndrome  
 CC and syndrome X.

SQ Sequence 43 AA;  
 Query Match 40.0%; Score 169; DB 21; Length 43;  
 Best Local Similarity 76.7%; Pred. No. 2e-14;

Matches 33; Conservative 4; Mismatches 6; Indels 0; Gaps 0;  
 OY 13 PMRSISENSLVAMDFSGKSRVIENPTEALSVAVEEGLAWRK 55  
 DB 1 PVRSVSENSLVAMDFSGTGRVIENPTEALSVALEEAQAWRK 43

RESULT 11  
 AABI8957  
 ID AABI8957 standard; peptide: 43 AA.

XX AC AABI8957;  
 XX DT 08-FEB-2001 (first entry)  
 XX DE Peptide derived from the PIR or PIR-SH2 domain of Grb7 protein.  
 XX KW Phosphorylated insulin receptor interacting region; Grb7 family protein;  
 KW insulin receptor; tyrosine kinase; insulin; insulin-associated disease;  
 KW diabetes; obesity; polycystic ovarian syndrome; syndrome X.  
 XX OS Homo sapiens.  
 XX PN WO200055634-A1.  
 XX PD 21-SEP-2000.  
 XX PF 14-MAR-2000; 2000WO-FR00613.  
 XX PR 15-MAR-1999; 99FR-0003159.  
 XX PA (CNRS ) CNRS CENT NAT RECH SCI.  
 XX PI Burnol A, Perdureau D, Kasus-Jacobi A, Bereziat V, Girard J;  
 DR WPI; 2000-587566/55.  
 XX PT Fragments of Grb family proteins to identify compounds are useful in  
 PT treating insulin-associated diseases, particularly diabetes and obesity  
 PT -

Claim 2; Page 34; 46pp; French.

XX BI8937-64 represent the PIR (phosphorylated insulin receptor interacting  
 CC region) or PIR-SH2 (Src homology 2) domains of a Grb7 family protein.  
 CC PIR is the actual binding region but its effect is about 10 times  
 CC greater in presence of SH2 (which by itself is inactive). Agents that  
 CC affect binding between the peptides and the insulin receptor can  
 CC stimulate or inhibit tyrosine kinase activity of the receptor. The  
 CC peptides are used for screening molecules for ability to treat diseases  
 CC in which insulin is implicated. The peptides are used to identify agents  
 CC that are potentially useful for treating insulin-associated diseases,  
 CC particularly diabetes and obesity but also polycystic ovarian syndrome  
 CC and syndrome X.

SQ Sequence 43 AA;

Query Match 38.3%; Score 162; DB 21; Length 43;  
 Best Local Similarity 74.4%; Pred. No. 1.6e-13;  
 Matches 32; Conservative 4; Mismatches 7; Indels 0; Gaps 0;  
 OY 13 PMRSISENSLVAMDFSGKSRVIENPTEALSVAVEEGLAWRK 55  
 DB 1 PLRSASDNTLVAMDFSGHAGRVENPREALSVALEEAQAWRK 43

RESULT 12  
 AABI8945  
 ID AABI8945 standard; peptide: 43 AA.

XX AC AABI8945;  
 XX DT 08-FEB-2001 (first entry)

XX DE Peptide derived from the PIR or PIR-SH2 domain of Grb7 protein.  
 XX KW Phosphorylated insulin receptor interacting region; Grb7 family protein;  
 XX KW insulin receptor; tyrosine kinase; insulin; insulin-associated disease;  
 XX KW diabetes; obesity; polycystic ovarian syndrome; syndrome X.  
 XX OS Mus muris.  
 XX PN WO200055634-A1.  
 XX PD 21-SEP-2000.  
 XX PF 14-MAR-2000; 2000WO-FR00613.  
 XX PR 15-MAR-1999; 99FR-0003159.  
 XX PA (CNRS ) CNRS CENT NAT RECH SCI.  
 XX PI Burnol A, Perdureau D, Kasus-Jacobi A, Bereziat V, Girard J;  
 XX DR WPI: 2000-587566/55.  
 XX XX Fragments of Grb family proteins to identify compounds are useful in  
 PT treating insulin-associated diseases, particularly diabetes and obesity  
 PT  
 XX PS Claim 2; Page 27-28; 46pp; French.  
 XX CC BL8937-64 represent the PIR (phosphorylated insulin receptor interacting  
 CC region) or PIR-SH2 (Src homology 2) domains of a Grb7 family protein.  
 CC PIR is the actual binding region but its effect is about 10 times  
 CC greater in presence of SH2 (which by itself is inactive). Agents that  
 CC affect binding between the peptides and the insulin receptor can  
 CC stimulate or inhibit tyrosine kinase activity of the receptor. The  
 CC peptides are used for screening molecules for ability to treat diseases  
 CC in which insulin is implicated. The peptides are used to identify agents  
 CC that are potentially useful for treating insulin-associated diseases,  
 CC particularly diabetes and obesity but also polycystic ovarian syndrome  
 CC and syndrome X.  
 XX SQ Sequence 43 AA;  
 XX Query Match 38.1%; Score 161; DB 21; Length 43;  
 XX Best Local Similarity 78.0%; Pred. No. 2.2e-13;  
 XX Matches 32; Conservative 3; Mismatches 6; Indels 0; Gaps 0;  
 QY 13 PMRSISENSLVAMDFSGQKSRVIENPTREALSVAVEEGLAWR 53  
 DB 1 PMRSISENSLVAMDFSGQIGRVIDNPRAEQSALEBHW 41  
 RESULT 13  
 AAB18953  
 ID AAB18953 standard; peptide; 43 AA.  
 AC AAB18953;  
 DT 08-FEB-2001 (first entry)  
 DE Peptide derived from the PIR or PIR-SH2 domain of Grb7 protein.  
 KW Phosphorylated insulin receptor interacting region; Grb7 family protein;  
 KW insulin receptor; tyrosine kinase; insulin; insulin-associated disease;  
 KW diabetes; obesity; polycystic ovarian syndrome; syndrome X.  
 XX Rattus sp.  
 XX PN WO200055634-A1.  
 XX PD 21-SEP-2000.  
 XX PF 14-MAR-2000; 2000WO-FR00613.

XX PR 15-MAR-1999; 99FR-0003159.  
 XX PA (CNRS ) CNRS CENT NAT RECH SCI.  
 XX PI Burnol A, Perdureau D, Kasus-Jacobi A, Bereziat V, Girard J;  
 XX DR WPI: 2000-587566/55.  
 XX XX Fragments of Grb family proteins to identify compounds are useful in  
 PT treating insulin-associated diseases, particularly diabetes and obesity  
 PT  
 XX PS Claim 2; Page 32; 46pp; French.  
 XX CC BL8937-64 represent the PIR (phosphorylated insulin receptor interacting  
 CC region) or PIR-SH2 (Src homology 2) domains of a Grb7 family protein.  
 CC PIR is the actual binding region but its effect is about 10 times  
 CC greater in presence of SH2 (which by itself is inactive). Agents that  
 CC affect binding between the peptides and the insulin receptor can  
 CC stimulate or inhibit tyrosine kinase activity of the receptor. The  
 CC peptides are used for screening molecules for ability to treat diseases  
 CC in which insulin is implicated. The peptides are used to identify agents  
 CC that are potentially useful for treating insulin-associated diseases,  
 CC particularly diabetes and obesity but also polycystic ovarian syndrome  
 CC and syndrome X.  
 XX SQ Sequence 43 AA;  
 XX Query Match 37.6%; Score 159; DB 21; Length 43;  
 XX Best Local Similarity 69.8%; Pred. No. 4e-13;  
 XX Matches 30; Conservative 6; Mismatches 7; Indels 0; Gaps 0;  
 QY 13 PMRSISENSLVAMDFSGQKSRVIENPTREALSVAVEEGLAWR 55  
 DB 1 PLRSVSDNTLVAMDFSGHAGRVIDNPREALSAAAEQAQRKK 43  
 RESULT 14  
 AAB18961  
 ID AAB18961 standard; peptide; 43 AA.  
 AC AAB18961;  
 DT 08-FEB-2001 (first entry)  
 DE Peptide derived from the PIR or PIR-SH2 domain of Grb7 protein.  
 KW Phosphorylated insulin receptor interacting region; Grb7 family protein;  
 KW insulin receptor; tyrosine kinase; insulin; insulin-associated disease;  
 KW diabetes; obesity; polycystic ovarian syndrome; syndrome X.  
 XX Mus muris.  
 XX PN WO200055634-A1.  
 XX PD 21-SEP-2000.  
 XX PF 14-MAR-2000; 2000WO-FR00613.  
 XX PR 15-MAR-1999; 99FR-0003159.  
 XX PA (CNRS ) CNRS CENT NAT RECH SCI.  
 XX PI Burnol A, Perdureau D, Kasus-Jacobi A, Bereziat V, Girard J;  
 XX DR WPI: 2000-587566/55.  
 XX XX Fragments of Grb family proteins to identify compounds are useful in  
 PT treating insulin-associated diseases, particularly diabetes and obesity  
 PT  
 XX PS Claim 2; Page 36; 46pp; French.



XX B18937-64 represent the PIR (phosphorylated insulin receptor interacting  
CC region) or PIR-SH2 (Src homology 2) domains of a Grb7 family protein.  
CC PIR is the actual binding region but its effect is about 10 times that  
CC greater in presence of SH2 (which by itself is inactive). Agents that  
CC affect binding between the peptides and the insulin receptor can  
CC stimulate or inhibit tyrosine kinase activity of the receptor. The  
CC peptides are used for screening molecules for ability to treat diseases  
CC in which insulin is implicated. The peptides are used to identify agents  
CC that are potentially useful for treating insulin-associated diseases,  
CC particularly diabetes and obesity but also polycystic ovarian syndrome  
CC and syndrome X.  
XX  
SQ Sequence 43 AA;

Query Match 37.6%; Score 159; DB 21; Length 43;  
Best Local Similarity 69.8%; Pred. No. 4e-13;  
Matches 30; Conservative 6; Mismatches 7; Indels 0; Gaps 0;

QY 13 PKRSISENSLVAMDFSGKSRVIENTEALSVAVEGLAWRK 55  
DB 1 PLRSVSDNTLVAMDFSGHAGRVIDNPREALSAAMEEAQAWRK 43

RESULT 15  
AAO02215  
ID AAO02215 standard; Protein; 47 AA.  
AC AAO02215;  
XX  
XX  
DT 06-NOV-2001 (first entry)  
XX  
DE Human polypeptide SEQ ID NO 16107.

XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
KW nervous system disorders; arthritis; inflammation.  
XX  
OS Homo sapiens.  
XX  
XX WO200164835-A2.  
XX  
XX 07-SEP-2001.  
XX  
XX 26-FEB-2001; 2001WO-US04927.  
XX  
XX 28-FEB-2000; 2000US-0515126.  
XX  
XX 18-MAY-2000; 2000US-0577409.  
XX  
XX (HYSE-) HYSEQ INC.  
XX  
XX Tang YT, Liu C, Drmanac RT;  
XX  
XX WPI: 2001-514838/56.  
XX  
XX N-PSDB; AAI82146.

XX Isolated nucleic acids and polypeptides, useful for preventing  
PT diagnosing and treating e.g. leukaemia, inflammation and immune  
PT disorders -  
XX  
XX Claim 20; SEQ ID NO 16107; 1399pp + Sequence Listing; English.  
XX  
XX The invention relates to human polynucleotides (AAI79941-AAI93841) and  
CC the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to  
CC cytokine, cell proliferation or cell differentiation or which may induce  
CC production of other cytokines in other cell populations. The  
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
CC peptide therapy. The polypeptides have various cytokine-like activities,  
CC e.g. stem cell growth factor activity, haematopoiesis regulating  
CC activity, tissue growth factor activity, immunomodulatory activity and  
CC activin/inhibin activity and may be useful in the diagnosis and/or  
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and

CC inflammation.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 47 AA;

Query Match 15.5%; Score 65.5; DB 22; Length 47;  
Best Local Similarity 38.5%; Pred. No. 0.6;  
Matches 15; Conservative 8; Mismatches 13; Indels 3; Gaps 1;

QY 48 EGLAWRKGCCLRLGRPHGS---PTASSOSSATNMAIHRSQ 83  
DB 4 DGVWRNPGSLQPPSGSDPPTSASOESGTTGAHHTR 42

RESULT 16  
AAO13559  
ID AAO13559 standard; Protein; 77 AA.  
AC AAO13559;  
XX  
XX 06-NOV-2001 (first entry)  
XX  
DE Human polypeptide SEQ ID NO 27451.

XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
KW nervous system disorders; arthritis; inflammation.  
XX  
OS Homo sapiens.  
XX  
XX WO200164835-A2.  
XX  
XX 07-SEP-2001.  
XX  
XX 26-FEB-2001; 2001WO-US04927.  
XX  
XX 28-FEB-2000; 2000US-0515126.  
XX  
XX 18-MAY-2000; 2000US-0577409.  
XX  
XX (HYSE-) HYSEQ INC.  
XX  
XX Tang YT, Liu C, Drmanac RT;  
XX  
XX WPI: 2001-514838/56.  
XX  
XX N-PSDB; AAI93490.

XX Isolated nucleic acids and polypeptides, useful for preventing  
PT diagnosing and treating e.g. leukaemia, inflammation and immune  
PT disorders -  
XX  
XX Claim 20; SEQ ID NO 27451; 1399pp + Sequence Listing; English.  
XX  
XX The invention relates to human polynucleotides (AAI79941-AAI93841) and  
CC the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to  
CC cytokine, cell proliferation or cell differentiation or which may induce  
CC production of other cytokines in other cell populations. The  
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
CC peptide therapy. The polypeptides have various cytokine-like activities,  
CC e.g. stem cell growth factor activity, haematopoiesis regulating  
CC activity, tissue growth factor activity, immunomodulatory activity and  
CC activin/inhibin activity and may be useful in the diagnosis and/or  
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and

XX Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 77 AA;

Query Match 14.4%; Score 61; DB 22; Length 77;  
Best Local Similarity 38.5%; Pred. No. 0.6;  
Matches 15; Conservative 8; Mismatches 13; Indels 3; Gaps 1;



[illegible]

PD 25-OCT-2001.  
 XX 16-APR-2001; 2001WO-US08656.  
 XX 18-APR-2000; 2000US-0552929.  
 PR 26-JAN-2001; 2001US-0770160.  
 XX (HYSE-) HYSEQ INC.  
 PA Tang YT, Liu C, Drmanac RT;  
 PI WPI; 2001-611725/70.  
 DR Nucleic acids encoding a range of human polypeptides, useful in genetic  
 PT vaccination, testing and therapy -  
 XX Claim 20; Page 366; 765pp; English.  
 XX The invention relates to novel human secreted polypeptides. The  
 CC polypeptides and antibodies to the polypeptides are useful for  
 CC determining the presence of or predisposition to a disease associated  
 CC with altered levels of polypeptide. The polypeptides are also useful for  
 CC identifying agents (agonists and antagonists) that bind to them. Cells  
 CC expressing the proteins are useful for identifying a therapeutic agent  
 CC for use in treatment of a pathology related to aberrant expression or  
 CC physiological interactions of the polypeptides. Vectors comprising  
 CC the nucleic acids encoding the polypeptides and cells genetically  
 CC engineered to express them are also useful for producing the proteins.  
 CC The proteins are useful in genetic vaccination, testing and  
 CC therapy, and can be used as nutritional supplements. They may be used to  
 CC increase stem cell proliferation; to regulate haematopoiesis; and in  
 CC bone, cartilage, tendon and/or nerve tissue growth or regeneration;  
 CC immune suppression and/or stimulation; as anti-inflammatory agents; and  
 CC in treatment of leukaemias. AAU29510-AAU33304 represent the amino acid  
 CC sequences of novel human secreted proteins of the invention.  
 XX Sequence 72 AA;  
 SQ

Query Match 13.2%; Score 56; DB 22; Length 72;  
 Best Local Similarity 27.9%; Pred. No. 19;  
 Matches 17; Conservative 14; Mismatches 16; Indels 14; Gaps 3;  
 OY 8 SQSIPMRISSEISLVAMDFSGKSRV-----IENPTEALSV-----VBEGLAWRK 54  
 DB 11 SSSNPLSSXLLNKIPSLPSSWEXWIPPRNNCLSLNLPSP-SUAPSLDDIKEGLSWKR 69  
 OY 55 K 55  
 DB 70 K 70

RESULT 22  
 ABG65212  
 ID ABG65212 standard; Protein; 79 AA.  
 XX AC ABG65212;  
 XX DT 27-AUG-2002 (first entry)  
 XX DE Human albumin fusion protein #1887.  
 KW Albumin fusion protein; therapeutic protein X; human albumin; HA;  
 KW human serum albumin; HSA; cancer; reproductive disorder;  
 KW digestive disorder; immune disorder; endocrine disorder;  
 KW haematopoietic disorder; neural disorder; connective disorder;  
 KW cytostatic; antifertility; antiinflammatory; antiulcer;  
 KW immunomodulator; anti-HIV; antidiabetic; haemostatic; neutropic;  
 KW neuroprotective; antiparkinsonian; antimicrobial; neuroleptic;  
 KW osteopathic; antiarthritis.  
 XX OS Homo sapiens.  
 OS Synthetic.

PN WO200177137-A1.  
 XX PD 18-OCT-2001.  
 XX PF 12-APR-2001; 2001WO-US11988.  
 XX PR 12-APR-2000; 2000US-229358P.  
 PR 25-APR-2000; 2000US-199384P.  
 PR 21-DEC-2000; 2000US-256931P.  
 XX (HUNA-) HUMAN GENOME SCI INC.  
 XX Rosen CA, Haseltine WA;  
 PI WPI; 2002-010886/01.  
 DR New fusion protein for treating disease e.g. diabetes comprises an  
 PT albumin fused to a therapeutic protein -  
 XX Claim 1; Page 1828-1829; 2102pp; English.  
 XX The present invention relates to albumin fusion proteins comprising a  
 CC therapeutic protein X and human albumin (HA, also known as human serum  
 CC albumin, HSA). The proteins are useful for treating a disease or  
 CC disorder that may be modulated by therapeutic protein X. The albumin  
 CC extends the shelf-life of protein X, and may increase its biological  
 CC in vitro/in vivo activity. The protein is useful for treating and  
 CC diagnosing disorders such as cancer, reproductive disorders, digestive  
 CC disorders (e.g. Crohn's disease, ulcerative colitis), immune disorders  
 CC (e.g. acquired immunodeficiency syndrome, AIDS), endocrine disorders  
 CC (e.g. diabetes), haematopoietic disorders, neural disorders  
 CC (e.g. Alzheimer's, Parkinson's, Creutzfeldt-Jacob disease,  
 CC encephalomyelitis, meningitis, schizophrenia), and connective disorders  
 CC (e.g. osteoporosis, arthritis). ABG63326-ABG65518 represent albumin  
 CC fusion proteins of the invention.  
 XX Sequence 79 AA;  
 SQ

Query Match 13.2%; Score 56; DB 23; Length 79;  
 Best Local Similarity 39.3%; Pred. No. 22;  
 Matches 11; Conservative 3; Mismatches 14; Indels 0; Gaps 0;  
 OY 57 CLRGLTGHTASSQSSATNMAIHRSQP 84  
 DB 45 CLSIGQHELPSYSCQPGKRKLPHHSQP 72

RESULT 23  
 AAU91124  
 ID AAU91124 standard; Protein; 79 AA.  
 XX AC AAU91124;  
 XX DT 05-JUN-2002 (first entry)  
 XX DE Human secreted protein sequence #44.  
 KW Human secreted protein; autoimmune disease; hyperproliferative disorder;  
 KW cardiovascular disorder; cerebrovascular disorder; infection; cancer;  
 KW nervous system disorder; ocular disorder; epithelial cell proliferation;  
 KW wound healing; skin aging; sunburn; transplantation; chemotaxis;  
 KW tissue regeneration; food additive; preservative; cytostatic; cardiant;  
 KW antiviral; antiallergic; antiinflammatory; antibacterial; antifungal.  
 XX OS Homo sapiens.  
 XX WO200218412-A1.  
 XX PD 07-MAR-2002.  
 XX PF 17-JAN-2001; 2001WO-US01384.  
 XX PR 28-AUG-2000; 2000US-228086P.

PR 04-JAN-2001; 2001US-259516P.  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX Rosen CA, Komatsoulis GA, Baker KP, Birse CE, Soppet DR, Olsen HS;  
PI Moore PA, Wei P, Ebner R, Duan RD, Shi Y, Choi GH, Fiscella M;  
PI Ni J;  
XX WPI; 2002-269525/31.  
DR N-PSDB; ABK54162.  
XX Seventeen nucleic acid molecules encoding human secreted proteins,  
PT useful in the prevention, treatment and diagnosis of cancer, immune  
PT disorders, cardiovascular disorders and neurological diseases -  
XX Claim 11; Page 478-479; 505pp; English.  
XX The present invention relates to the isolation of novel human secreted  
CC proteins' and the polynucleotide sequences encoding them. The secreted  
CC proteins are useful to prevent, treat or ameliorate a medical condition  
CC in e.g. humans, mice, rabbits, goats, horses, cats, dogs, chickens or  
CC sheep. The secreted proteins are also useful in diagnosing a pathological  
CC condition or susceptibility to a pathological condition. Antibodies to  
CC the secreted proteins can also be used in alleviating symptoms associated  
CC with disorders and in diagnostic immunoassays e.g. radioimmunoassays or  
CC enzyme linked immunosorbent assays (ELISA). Disorders which can be  
CC diagnosed or treated include autoimmune diseases e.g. rheumatoid  
CC arthritis, hyperproliferative disorders e.g. cancer, cardiovascular  
CC disorders e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral  
CC ischaemia, angiodysplasia, nervous system disorders e.g. Parkinson's  
CC disease, infections caused by bacteria, viruses and fungi and  
CC ocular disorders e.g. corneal infection. The polypeptides can also be  
CC used to aid wound healing and epithelial cell proliferation, to prevent  
CC skin aging due to sunburn, to maintain organs before transplantation, for  
CC supporting cell culture of primary tissues, to regenerate tissues and in  
CC chemotaxis. The polypeptides can also be used as a food additive or  
CC preservative to increase or decrease storage capabilities.  
CC AAU91081-AAU91148 represent human secreted protein sequences.  
XX Sequence 79 AA:  
SQ Query Match 13.2%; Score 56; DB 23; Length 79;  
Best Local Similarity 39.3%; Pred. No. 22;  
Matches 11; Conservative 3; Mismatches 14; Indels 0; Gaps 0;  
QY 57 CLRUGTSGPTASSQSSATNMATHRSQP 84  
DB 45 CLSIGQHELPSYSCQPGKRLPHHSQP 72  
RESULT 24  
AAO00883  
ID AAO00883 standard; Protein; 49 AA.  
XX AAO00883;  
AC AAO00883;  
XX 06-NOV-2001 (first entry)  
DT Human polypeptide SEQ ID NO 14775.  
DE Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
XX Human vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
KW nervous system disorders; arthritis; inflammation.  
XX Homo sapiens.  
OS WO200164835-A2.  
PN 07-SEP-2001.  
PD 26-FEB-2001; 2001WO-US04927.  
PR 28-FEB-2000; 2000US-0515126.  
XX 18-MAY-2000; 2000US-0577409.  
XX (HYSE-) HYSEQ INC.  
Tang YT, Liu C, Drmanac RT;  
WPI; 2001-514838/56.

PR 28-FEB-2000; 2000US-0515126.  
XX 18-MAY-2000; 2000US-0577409.  
XX (HYSE-) HYSEQ INC.  
Tang YT, Liu C, Drmanac RT;  
WPI; 2001-514838/56.  
DR N-PSDB; AAI80814.  
XX Isolated nucleic acids and polypeptides, useful for preventing  
PT diagnosing and treating e.g. leukaemia, inflammation and immune  
PT disorders -  
XX Claim 20; SEQ ID NO 14775; 1399pp + Sequence Listing; English.  
XX The invention relates to human polynucleotides (AAI79941-AAI93841) and  
CC the encoded proteins (AAO0010-AAO13910) that exhibit activity relating to  
CC cytokine, cell proliferation or cell differentiation or which may induce  
CC production of other cytokines in other cell populations. The  
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
CC peptide therapy. The polypeptides have various cytokine-like activities,  
CC e.g. stem cell growth factor activity, haematopoiesis regulating  
CC activity, tissue growth factor activity, immunomodulatory activity and  
CC activin/inhibin activity and may be useful in the diagnosis and/or  
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
CC inflammation.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX Sequence 49 AA:  
SQ Query Match 13.1%; Score 55.5; DB 22; Length 49;  
Best Local Similarity 33.3%; Pred. No. 13;  
Matches 13; Conservative 8; Mismatches 15; Indels 3; Gaps 1;  
QY 48 EGLAWKKGLRLGTHGS---PTASSQSSATNMATHRSQ 83  
DB 6 DGVPRNFGSLKPPSGSDPTISASOECGITGAHHHTR 44  
RESULT 25  
AAO06915  
ID AAO06915 standard; Protein; 73 AA.  
XX AAO06915;  
AC AAO06915;  
XX 06-NOV-2001 (first entry)  
DT Human polypeptide SEQ ID NO 20807.  
DE Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
XX Human vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
KW nervous system disorders; arthritis; inflammation.  
XX Homo sapiens.  
OS WO200164835-A2.  
PN 07-SEP-2001.  
PD 26-FEB-2001; 2001WO-US04927.  
PR 28-FEB-2000; 2000US-0515126.  
XX 18-MAY-2000; 2000US-0577409.  
XX (HYSE-) HYSEQ INC.  
Tang YT, Liu C, Drmanac RT;  
WPI; 2001-514838/56.

DR N-PSDB: AAI86846.  
 XX Isolated nucleic acids and polypeptides, useful for preventing  
 PT diagnosing and treating e.g. leukaemia, inflammation and immune  
 PT disorders -  
 XX  
 PS  
 PS Claim 20: SEQ ID NO 20807; 1399pp + Sequence Listing; English.  
 XX  
 CC The invention relates to human polynucleotides (AAI79941-AAI93841) and  
 CC the encoded proteins (AAO00010-AAO13910) that exhibit activity relating to  
 CC cytokine, cell proliferation or cell differentiation or which may induce  
 CC production of other cytokines in other cell populations. The  
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
 CC peptide therapy. The polypeptides have various cytokine-like activities,  
 CC e.g. stem cell growth factor activity, haematopoiesis regulating  
 CC activity, tissue growth factor activity, immunomodulatory activity and  
 CC activin/inhibin activity and may be useful in the diagnosis and/or  
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
 CC inflammation.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 73 AA:

Query Match 13.0%; Score 55; DB 22; Length 73;  
 Best Local Similarity 44.4%; Pred. NO. 26;  
 Matches 12; Conservative 4; Mismatches 11; Indels 0; Gaps 0;

QY 58 LRLGTHGSPATASOSSATNMAIHRSP 84  
 ||||| |::|:| |  
 Db 23 LRLGLSDPPASASESTGTGMSHCSP 49

Search completed: March 28, 2003, 09:05:51  
 Job time : 87.1417 secs

GenCore version 5.1.4.p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 28, 2003, 09:01:30 ; Search time 35,2126 seconds  
(without alignments)  
251.615 Million cell updates/sec

Title: us-09-936-697-5  
Effect score: -212  
Sequence: 1 PMSISENSILVAMDFSGOKS.....ENPTEALSVAVEGLAWKKK 43

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 74486

Minimum DB seq length: 40  
Maximum DB seq length: 90

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 100 summaries

Database: 1: sp\_archaea: \*  
2: sp\_bacteria: \*  
3: sp\_fungi: \*  
4: sp\_human: \*  
5: sp\_invertebrate: \*  
6: sp\_mammal: \*  
7: sp\_mhc: \*  
8: sp\_organelle: \*  
9: sp\_phase: \*  
10: sp\_plant: \*  
11: sp\_rodent: \*  
12: sp\_virus: \*  
13: sp\_vertebrate: \*  
14: sp\_unclassified: \*  
15: sp\_rvirus: \*  
16: sp\_bacteriap: \*  
17: sp\_archaeap: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	45.5	21.5	73	5	P91302 caenorhabdi
2	45	21.2	77	17	O28902 archaeoglob
3	45	21.2	82	17	O27686 methanobact
4	44	20.8	60	17	O8ttk0 methanosarc
5	44	20.8	79	2	O9rcd4 xanthomonas
6	44	20.8	82	5	O9vxc1 drosophila
7	43.5	20.5	69	15	O9wmg6 human immun
8	43.5	20.5	88	10	O8lm11 arabidopsis
9	42.5	20.0	80	3	O9hgr8 choanephora
10	42	19.8	53	16	O9pfg5 xyliella fas
11	42	19.8	58	9	O80316 bacterioph
12	42	19.8	76	12	O64944 avian infec
13	42	19.8	76	12	O64947 avian infec
14	42	19.8	79	16	O328x5 chlamydia p
15	42	19.8	79	16	O9js88 chlamydia p
16	42	19.8	81	16	O9k247 chlamydia p

17	42	19.8	90	2	P72094
18	41.5	19.6	87	10	O40019
19	41	19.3	69	10	O41693
20	41	19.3	79	4	O9nrp2 homo sapien
21	41	19.3	85	2	O9wkg1 xanthomonas
22	41	19.3	85	16	O8s9d7 thermoanaer
23	41	19.3	87	17	O8rkz4 methanosarc
24	40.5	19.1	68	15	O74620 human immun
25	40.5	19.1	68	15	O74630 human immun
26	40.5	19.1	69	15	O9wmg5 human immun
27	40.5	19.1	79	2	O936t5 pseudomonas
28	40.5	19.1	79	2	O936t5 pseudomonas
29	40.5	19.1	88	16	O98kz6 rhizobium l
30	40.5	19.1	89	15	O9yv28 human immun
31	40	18.9	67	16	O97is7 clostridium
32	40	18.9	70	16	P71998 mycobacteri
33	40	18.9	73	3	O02288 coccidioid
34	40	18.9	84	12	O9wkh1 encephalomy
35	40	18.9	89	2	O9akh7 rickettsia
36	39.5	18.6	56	2	O9kk61 mycobacteri
37	39.5	18.6	58	5	O27193 tetrahymena
38	39.5	18.6	61	16	O9chz4 lactococcus
39	39.5	18.6	67	16	O92em1 listeria in
40	39.5	18.6	67	16	O910t9 streptomyce
41	39.5	18.6	68	15	O69653 human immun
42	39.5	18.6	69	15	O9wmg3 human immun
43	39.5	18.6	73	4	O9B2L1 homo sapien
44	39.5	18.6	73	11	O9EPV8
45	39.5	18.6	77	16	O8yxu8 anabaena sp
46	39.5	18.6	80	15	O9gst4 human immun
47	39.5	18.6	82	10	O9LNN9 arabidopsis
48	39.5	18.6	89	10	O22414 oryza sativ
49	39.5	18.6	89	15	O9yv38 human immun
50	39.5	18.6	89	15	O9yv37 human immun
51	39.5	18.6	89	15	O9yv35 human immun
52	39.5	18.6	89	15	O9yv25 human immun
53	39.5	18.6	89	15	O9yv24 human immun
54	39.5	18.6	89	15	O9wA05 human immun
55	39.5	18.6	89	15	O9w8F1 human immun
56	39.5	18.6	89	15	O9yjk2 human immun
57	39.5	18.6	89	15	O9yJ41 human immun
58	39.5	18.6	89	15	O9w9S7 human immun
59	39.5	18.6	89	15	O9w8S2 human immun
60	39	18.4	41	6	O18852 macaca radi
61	39	18.4	52	4	O96GJ3 homo sapien
62	39	18.4	54	10	O8YV75 arabidopsis
63	39	18.4	62	10	O9T0H2 arabidopsis
64	39	18.4	67	4	O9HIL3 homo sapien
65	39	18.4	68	6	P79120 bos taurus
66	39	18.4	72	16	O8YDJ3 brucella me
67	39	18.4	73	12	O9YPL4 encephalomy
68	39	18.4	76	12	O64948 avian infec
69	39	18.4	77	16	O9hzJ7 pseudomonas
70	39	18.4	77	16	O9k4K2 streptomyce
71	39	18.4	80	12	O91726 grapevine r
72	39	18.4	80	12	O91903 rupestris s
73	39	18.4	80	16	O9PHK1 xyliella fas
74	39	18.4	90	16	O92854 chlamydia p
75	38.5	18.2	67	16	O9x7N4 streptomyce
76	38.5	18.2	67	16	O8Y9V4 listeria mo
77	38.5	18.2	67	16	O8Y4M8 listeria mo
78	38.5	18.2	68	15	O74093 human immun
79	38.5	18.2	68	15	O74645 human immun
80	38.5	18.2	69	2	O51902 proteus mlr
81	38.5	18.2	69	11	O9CWL5 mus musculus
82	38.5	18.2	69	15	O9wmt2 human immun
83	38.5	18.2	80	14	O9SIV0 uncultured
84	38.5	18.2	81	15	O900Q6 human immun
85	38.5	18.2	86	11	O99iv9 mus musculus
86	38.5	18.2	87	13	O90X98 meleagris g
87	38.5	18.2	89	10	O9LGN8 oryza sativ
88	38	17.9	53	5	O27909 drosophila
89	38	17.9	64	16	O92j17 rickettsia

















```

OS Methanobacterium thermoautotrophicum.
OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;
OC Methanobacteriaceae; Methanothermobacter.
OX NCBI_TaxID=187420;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DELTA H;
RX MEDLINE=98037514; PubMed=9371463;
RA Smith D.R., Doucette-Stamm L.A., DeLonghery C., Lee H.-M., Dubois J.,
RA Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
RA Harrison D., Hoang L., Keagle P., Lum W., Pothier B., Qiu D.,
RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
RA Jiwani N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
RA McDougall S., Shiner G., Goyal A., Pietrowski S., Church G.M.,
RA Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.,
RA "Complete genome sequence of Methanobacterium thermoautotrophicum
RT deltaH; functional analysis and comparative genomics.";
RL J. Bacteriol. 179:7135-7155(1997).
DR EMBL: AE000924; AAB86122.1;
DR InterPro: IPR001109; HupF_HupC.
DR Pfam: PF01455; HupF_HupC; 1.
DR PRODOM: PD003112; HupF_HupC; 1.
DR TIGRFXMS: TIGR00074; hupC_hupF; 1.
KW Complete proteome.
SQ SEQUENCE 82 AA; 9082 MW; B6EGAED010FB6E2D CRC64;

Query Match 21.2%; Score 45; DB 17; Length 82;
Best Local Similarity 28.9%; Pred. No. 1.6e+02;
Matches 11; Conservative 9; Mismatches 14; Indels 4; Gaps 1;

QY 6 SENSILVDFSGOKSRV----IENPTEALSVAVEGLA 39
DB 14 SEDNIATVDFGGVROQKLDLVDVDEEGKYLVLHSGVA 51

RESULT 4
Q8TTK0 PRELIMINARY; PRT; 60 AA.
AC Q8TTK0;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DE Ferredoxin.
GN MA0431.
OS Methanosarcina acetivorans.
OC Archaea; Euryarchaeota; Methanococci; Methanosarcinales;
OC Methanosarcinaceae; Methanosarcina.
OX NCBI_TaxID=2214;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C2A / ATCC 35395 / DSM 2834;
RX MEDLINE=21929760; PubMed=11932238;
RA Galagan J.E., Nusbaum C., Roy A., Endrizzi M.G., MacDonald P.,
RA FitzHugh W., Calvo S., Engels R., Smirnov S., Atncor D., Brown A.,
RA Allen N., Naylor J., Stange-Thomann N., DeArellano K., Johnson R.,
RA Linton L., McKean P., McKernan K., Talamas J., Tirrell A., Ye W.,
RA Zimmer A., Barber R.D., Cann I., Graham D.E., Grahame D.A., Guss A.M.,
RA Hedderich R., Ingram-Smith C., Kuetner H.C., Krzycki J.A.,
RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reese J.N., Smith K.,
RA Springer T.A., Unayam L.A., White O., White R.H., de Macario E.C.,
RA Ferry J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I.,
RA Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
RA Metcalf W.W., Birren B.,
RA "The genome of Methanosarcina acetivorans reveals extensive metabolic
RT and physiological diversity.";
RL Genome Res. 12:332-342(2002).
DR EMBL: AE010703; AM03878.1;
KW Complete proteome.
SQ SEQUENCE 60 AA; 6265 MW; 6D75EBDB4460C21F CRC64;

Query Match 20.8%; Score 44; DB 17; Length 60;
Best Local Similarity 22.8%; Pred. No. 1.5e+02;
Matches 12; Conservative 7; Mismatches 9; Indels 0; Gaps 0;

QY 13 MDFSQKSRVIE-NPTEA-LSVAVEGLAWRRK 43
DB 46 LELSAEQAKAVNAHLSEALTDVDEALAWASK 78

RESULT 6
Q9VXC1 PRELIMINARY; PRT; 82 AA.
AC Q9VXC1;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE CG9644 protein.
GN CG9644.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blaisei R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,

```





```

RESULT 9
O9HGR8 PRELIMINARY; PRT; 80 AA.
AC O9HGR8;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Glyceraldehyde-3-phosphate dehydrogenase (EC 1.2.1.12)
DE (Fragment).
DE GPD.
GN GPD.
OS Choanophora infundibulifera.
OC Eukaryota; Fungi; Zygomycota; Zygomycetes; Mucorales; Choanephoraceae;
OC Choanephora.
OX NCBI_TaxID=127959;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NRRL 2560;
RA Tamas P.;
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=NRRL 2560;
RA Papp T., Vastag M., Acs K., Vagvolgyi C.;
RT Phylogenetic relationships among Mucoraceae, Choanephoraceae and
RT Gilbertellaceae based on rDNA and glyceraldehyde-3-phosphate
RT dehydrogenase sequences.
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: D-GLYCERALDEHYDE 3-PHOSPHATE +
CC NAD(+) = 3-PHOSPHO-D-GLYCEROL PHOSPHATE + NADH.
CC -1- SIMILARITY: BELONGS TO THE GLYCERALDEHYDE 3-PHOSPHATE
CC DEHYDROGENASE FAMILY.
DR EMBL: AJ278315; CAC05662.1;
DR InterPro: IPR000173; GAP_dhdrogenase.
DR Pfam: PF02800; gpdh_C.1.
KW Glycolysis; NAD; Oxidoreductase.
FT NON_TER 1
FT NON_TER 80
SQ SEQUENCE 80 AA; 8504 MW; 4ECCBEAE035943D0 CRC64;

Query Match 20.0%; Score 42.5; DB 3; Length 80;
Best Local Similarity 42.9%; Pred. No. 3.4e+02;
Matches 9; Conservative 6; Mismatches 3; Indels 3; Gaps 1;

OY 1 PMSI---SENSLVAMDFSGQ 18
DB 37 PMKGILGYTENAVVSTDFGE 57

RESULT 10
O9PFG5 PRELIMINARY; PRT; 53 AA.
AC O9PFG5;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Hypothetical protein xF0694.
GN xF0694.
OS Xylella fastidiosa.
OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
OX Xylella.
OX NCBI_TaxID=2371;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=9A5C;
RX MEDLINE=20365717; PubMed=10910347;
RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
RA Alvaranga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrier H.,
RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,
RA Facinani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,

```

```

RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA HO P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA Machado M.A., Madeira A.B.N., Madeira H.M.F., Marino C.L.,
RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA da Silva J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
RA de Souza A.P., Terenzi M.F., Truffi D., Tsal S.M., Tsubako M.H.,
RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
RT "The genome sequence of the plant pathogen Xylella fastidiosa."
RL Nature 406:151-159(2000).
DR EMBL: AE003912; AAF83504.1;
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 53 AA; 5958 MW; 4B14AF832900832B CRC64;

Query Match 19.8%; Score 42; DB 16; Length 53;
Best Local Similarity 66.7%; Pred. No. 2.4e+02;
Matches 8; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 30 LSVAVEEGLAWR 41
DB 21 LGVGVGYAWR 32

RESULT 11
O80316 PRELIMINARY; PRT; 58 AA.
ID O80316;
AC O80316;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE OF52 (Fragment).
GN H.
OS Bacteriophage 186.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Myoviridae;
OC P2-like viruses.
OX NCBI_TaxID=29252;
RN [1]
RP SEQUENCE FROM N.A.
RA Xue Q.;
RT "Studies on the tail region of the temperate coliphage 186 genome."
RL Thesis (1993), University of Adelaide.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=98371265; PubMed=9705261;
RA Portelli R., Dodd I.B., Xue Q., Egan J.B.;
RT "The late-expressed region of the temperate coliphage 186 genome."
RL Virology 248:117-130(1998).
DR EMBL: U32222; AAC34169.1;
FT NON_TER 1
FT VARIANT 15 15 S -> *
FT VARIANT 51 51 Q -> *
SQ SEQUENCE 58 AA; 6491 MW; 1199113D8CDEB8E6 CRC64;

Query Match 19.8%; Score 42; DB 9; Length 58;
Best Local Similarity 38.9%; Pred. No. 2.7e+02;
Matches 7; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

OY 26 PTEALSVAVEEGLAWRK 43
DB 31 PSELYSLSLTELITWREK 48

```

## RESULT 12

Q64944 PRELIMINARY; PRT; 76 AA.  
ID O64944  
AC O64944  
DT 01-NOV-1996 (TREMBlrel. 01, Created)  
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
DE Spike protein (Fragment).  
GN S1.  
OS Avian infectious bronchitis virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;  
OC Coronaviridae; Coronavirus.  
OX NCBI\_TaxID=11120;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=AL1960;  
RX MEDLINE=97049060; PubMed=8893790;  
RA Wang C.H., Tsai C.T.;  
RT "Genetic grouping for the isolates of avian infectious bronchitis  
RT virus in Taiwan."  
RL Arch. Virol. 141:1677-1688(1996).  
DR EMBL: U38678; AAB47436.1; -  
DR InterPro: IPR002551; Corona\_S1.  
DR Pfam: PF01600; Corona\_S1; 1.  
FT NON\_TER 1 76  
FT SEQUENCE 76 AA; 761 MW; 9DA97501A9CB4FD1 CRC64;  
Query Match 19.8%; Score 42; DB 12; Length 76;  
Best Local Similarity 31.8%; Pred. No. 3.7e+02;  
Matches 7; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

QY 21 RVIENTEALSVAVEEGLAWRK 42  
DB 36 RIVNASSIAMTPVPGOGMWSK 57

## RESULT 13

Q64947 PRELIMINARY; PRT; 76 AA.  
ID O64947  
AC O64947  
DT 01-NOV-1996 (TREMBlrel. 01, Created)  
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
DE Spike protein (Fragment).  
GN S1.  
OS Avian infectious bronchitis virus.  
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;  
OC Coronaviridae; Coronavirus.  
OX NCBI\_TaxID=11120;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=AL1955;  
RX MEDLINE=97049060; PubMed=8893790;  
RA Wang C.H., Tsai C.T.;  
RT "Genetic grouping for the isolates of avian infectious bronchitis  
RT virus in Taiwan."  
RL Arch. Virol. 141:1677-1688(1996).  
DR EMBL: U38681; AAB47439.1; -  
DR InterPro: IPR002551; Corona\_S1.  
DR Pfam: PF01600; Corona\_S1; 1.  
FT NON\_TER 1 76  
FT SEQUENCE 76 AA; 7903 MW; 271F114FD4078521 CRC64;  
Query Match 19.8%; Score 42; DB 12; Length 76;  
Best Local Similarity 36.4%; Pred. No. 3.7e+02;  
Matches 8; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

QY 21 RVIENTEALSVAVEEGLAWRK 42  
DB 36 RIVNASSIAMTPVPGOGMWSK 57

## RESULT 14

Q928X5 PRELIMINARY; PRT; 79 AA.  
ID Q928X5  
AC Q928X5  
DT 01-MAY-1999 (TREMBlrel. 10, Created)  
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)  
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)  
DE Hypothetical protein CPN0209.  
GN CPN0209.  
OS Chlamydia pneumoniae (Chlamydophila pneumoniae).  
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydophila.  
OX NCBI\_TaxID=83558;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CWL029;  
RX MEDLINE=99206606; PubMed=10192388;  
RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,  
RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;  
RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis."  
RL Nat. Genet. 21:385-389(1999).  
DR EMBL: AF001607; AAD18362.1; -  
DR Hypothetical protein; Complete proteome.  
KW SEQUENCE 79 AA; 9196 MW; 2813A36311D4A49A CRC64;  
Query Match 19.8%; Score 42; DB 16; Length 79;  
Best Local Similarity 33.3%; Pred. No. 3.9e+02;  
Matches 8; Conservative 7; Mismatches 9; Indels 0; Gaps 0;

QY 15 FSGOKSRVIENTEALSVAVEEGL 38  
DB 28 FQGRTRVIAITPAGLAIAEQNI 51

## RESULT 15

Q9JSH8 PRELIMINARY; PRT; 79 AA.  
ID Q9JSH8  
AC Q9JSH8  
DT 01-OCT-2000 (TREMBlrel. 15, Created)  
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)  
DE Hypothetical protein CPJ0209.  
GN CPJ0209.  
OS Chlamydia pneumoniae (Chlamydophila pneumoniae).  
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydophila.  
OX NCBI\_TaxID=83558;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=J118;  
RX MEDLINE=20330349; PubMed=10871362;  
RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,  
RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;  
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138  
RT from Japan and CWL029 from USA."  
RL Nucleic Acids Res. 28:2311-2314(2000).  
DR EMBL: AF002545; BAA98419.1; -  
KW Hypothetical protein.  
SQ SEQUENCE 79 AA; 9212 MW; C70CA36311C3AFF7 CRC64;  
Query Match 19.8%; Score 42; DB 16; Length 79;  
Best Local Similarity 33.3%; Pred. No. 3.9e+02;  
Matches 8; Conservative 7; Mismatches 9; Indels 0; Gaps 0;

QY 15 FSGOKSRVIENTEALSVAVEEGL 38  
DB 28 FQGRTRVIAITPAGLAIAEQNI 51

## RESULT 16

Q9K247 PRELIMINARY; PRT; 81 AA.  
ID Q9K247  
AC Q9K247  
DT 01-OCT-2000 (TREMBlrel. 15, Created)

DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DE 01-OCT-2001 (TREMBlrel. 18, Last annotation update)  
 GN Hypothetical protein CP0557.  
 OS Chlamydia pneumoniae (Chlamydia pneumoniae).  
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.   
 OX NCBI\_TaxID=83558;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=AR39;  
 RX MEDLINE=20150255; PubMed=10684935;  
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,  
 RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,  
 RA Linier K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,  
 RA Winn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,  
 RA Eisen J., Fraser C.M.;  
 RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia  
 RT pneumoniae AR39.";  
 RL Nucleic Acids Res. 28:1397-1406(2000).  
 DR EMBL; AE002214; AAP38377.1; -;  
 DR TIGR; CP0557; -;  
 KW Hypothetical protein.  
 SQ SEQUENCE 81 AA; 9455 MW; C6A6483DA44594C2 CRC64;  
 Query Match 19.8%; Score 42; DB 16; Length 81;  
 Best Local Similarity 33.3%; Pred. No. 4e+02; Mismatches 9; Indels 0; Gaps 0;  
 Matches 8; Conservative 7;  
 QY 15 FSGQKSRVNPTEALSVAVEGL 38  
 DB 30 FQGRTRVIAITPAGLAIAYEQNI 53  
 RESULT 17  
 P72094  
 ID P72094 PRELIMINARY; PRT; 90 AA.  
 AC P72094;  
 DT 01-FEB-1997 (TREMBlrel. 02, Created)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
 DE Transposase homolog (Fragment).  
 OS Neisseria meningitidis.  
 OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.  
 OX NCBI\_TaxID=487;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=22491;  
 RX MEDLINE=97008141; PubMed=8855317;  
 RA Tinsley C.R., Nassif X.;  
 RT "Analysis of the genetic differences between Neisseria meningitidis  
 RT and Neisseria gonorrhoeae: two closely related bacteria expressing two  
 RT different pathogenicities.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 93:11109-11114(1996).  
 DR EMBL; U58763; AAC44489.1; -;  
 FT NON-TER 90  
 FT NON-TER 90  
 SQ SEQUENCE 90 AA; 10285 MW; 80FAA15ABF1F53AF CRC64;  
 Query Match 19.8%; Score 42; DB 2; Length 90;  
 Best Local Similarity 41.7%; Pred. No. 4.6e+02;  
 Matches 10; Conservative 6; Mismatches 8; Indels 0; Gaps 0;  
 QY 13 MDFSQKSRVNPTEALSVAVEE 36  
 DB 2 LDYSSQKRVAYDIHDADHIVKD 25  
 RESULT 18  
 Q40019  
 ID Q40019 PRELIMINARY; PRT; 87 AA.  
 AC Q40019;  
 DT 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)

DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE B12D mRNA.  
 GN HVBI2D OR B12DGI.  
 OS Hordeum vulgare (Barley).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;  
 OC Triticeae; Hordeum.  
 OX NCBI\_TaxID=4513;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. BOMI;  
 RX MEDLINE=94236153; PubMed=8180622;  
 RA Aalen R.B., Opsahl-Ferstad H.G., Lindestad C., Olsen O.A.;  
 RT "Transcripts encoding an oleosin and a dormancy-related protein are  
 RT present in both the aleurone layer and the embryo of developing barley  
 RT (Hordeum vulgare L.) seeds.";  
 RL Plant J. 5:385-396(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. BOMI;  
 RX Steinum T.M., Berner H., Stacy R.A.P., Salehian Z., Aalen R.B.;  
 RT "Differential regulation of the barley (Hordeum vulgare) transcripts  
 RT B22E and B12D in mature aleurone layers.";  
 RL Physiol. Plantarum 102:337-345(1998).  
 DR EMBL; X76604; CAA54065.1; -;  
 DR EMBL; X09805; CAA70936.1; -;  
 SQ SEQUENCE 87 AA; 9648 MW; EECE117231E51046 CRC64;  
 Query Match 19.6%; Score 41.5; DB 10; Length 87;  
 Best Local Similarity 35.1%; Pred. No. 5.1e+02;  
 Matches 13; Conservative 7; Mismatches 16; Indels 1; Gaps 1;  
 QY 2 MRSISESLVAMDFSGKSRVNPTEALSVAVEGL 38  
 DB 29 LRNITGNPEVRVSKMGAGVLENHEGRRYA-EHGL 64  
 RESULT 19  
 Q41693  
 ID Q41693 PRELIMINARY; PRT; 69 AA.  
 AC Q41693;  
 DT 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
 DE 1-aminocyclopropane 1-carboxylate synthase (Fragment).  
 OS Vigna radiata.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC Eurosid 1; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Vigna.  
 OX NCBI\_TaxID=157791;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=HYPOCOTYL;  
 RX MEDLINE=92216056; PubMed=1558953;  
 RA Botella J.R., Schlagnhauser C.D., Arteca R.N., Phillips A.T.;  
 RT "Identification and characterization of three putative genes for 1-  
 RT aminocyclopropane-1-carboxylate synthase from etiolated mung bean  
 RT hypocotyl segments.";  
 RL Plant Mol. Biol. 18:793-797(1992).  
 DR EMBL; M80555; AAK38750.1; -;  
 DR HSSP; P37821; 1B8G. 1  
 FT NON-TER 69  
 FT NON-TER 69  
 SQ SEQUENCE 69 AA; 8057 MW; EAD4BDE7D47735D CRC64;  
 Query Match 19.3%; Score 41; DB 10; Length 69;  
 Best Local Similarity 28.9%; Pred. No. 4.6e+02;  
 Matches 11; Conservative 9; Mismatches 14; Indels 4; Gaps 2;  
 QY 5 ISESLVAMDFSGKSRVNPTEALSVAVEGLAWRK 42  
 DB 27 LAENPQLTDLI--QDWIMKNP--QASVCTPEGVKWK 60

```

RESULT 20
Q9NRP2 PRELIMINARY; PRT; 79 AA.
ID Q9NRP2
AC Q9NRP2
DT 01-OCT-2000 (TRENBLrel. 15, Created)
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
DE DC13.
DE DC13.
GN Homo sapiens (Human).
OS Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Gu Y., Peng Y., Li N., Gu W., Han Z., Fu G., Chen Z.;
RT "Novel genes expressed in human dentritic cell.";
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF201935; AAF86871.1;
SQ SEQUENCE 79 AA; 9460 MW; 783381BD5DAFB7AA CRC64;

Query Match 19.3%; Score 41; DB 4; Length 79;
Best Local Similarity 40.0%; Pred. No. 5.4e+02;
Matches 10; Conservative 5; Mismatches 6; Indels 4; Gaps 1;

QY 19 KSRVIENPTALSVAVVEGLAWKK 43
DB 49 KNEYVENTKSR-----EHGIAMKK 69

RESULT 21
Q9WVG1 PRELIMINARY; PRT; 85 AA.
ID Q9WVG1
AC Q9WVG1
DT 01-NOV-1999 (TRENBLrel. 12, Created)
DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)
DE HRPD6.
DE HRPD6.
GN Xanthomonas campestris (pv. vesicatoria).
OS Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
OC Xanthomonas;
OX NCBI_TaxID=341;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=75-3;
RX MEDLINE=98453137; PubMed=978181876;
RA Huguet E., Hahn K., Wengelnik K., Bonas U.;
RT "hpaA mutants of Xanthomonas campestris pv. vesicatoria are affected
RT in pathogenicity but retain the ability to induce host-specific
RT hypersensitive reaction.";
RL Mol. Microbiol. 29:1379-1390(1998).
DR EMBL; AF056246; AAD21324.1;
SQ SEQUENCE 85 AA; 9420 MW; 6180906387986E71 CRC64;

Query Match 19.3%; Score 41; DB 2; Length 85;
Best Local Similarity 44.4%; Pred. No. 5.8e+02;
Matches 8; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 12 AMDFSGOKSRVIENPTEA 29
DB 26 AMDLSGELRNVTALDA 43

RESULT 22
Q8R9D7 PRELIMINARY; PRT; 85 AA.
ID Q8R9D7
AC Q8R9D7;
DT 01-JUN-2002 (TRENBLrel. 21, Created)
DT 01-JUN-2002 (TRENBLrel. 21, Last sequence update)
DE 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
DE Hypothetical protein TTE1677.

GN Thermoanaerobacter tengcongensis.
OS Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Thermoanaerobacteriales; Thermoanaerobacteriaceae; Thermoanaerobacter.
OX NCBI_TaxID=119072;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=MB4T / JCM11007;
RX MEDLINE=21592816; PubMed=11997336;
RA Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,
RA Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,
RA Tan H., Chen R., Wang J., Yu J., Tang H.;
RT "A complete sequence of T. tengcongensis genome.";
RL Genome Res. 12:689-700(2002).
DR EMBL; AE011122; AAM24878.1;
RW Hypothetical protein; Complete proteome.
SQ SEQUENCE 85 AA; 9722 MW; 03DAGADE30EDE71C CRC64;

Query Match 19.3%; Score 41; DB 16; Length 85;
Best Local Similarity 30.0%; Pred. No. 5.8e+02;
Matches 9; Conservative 9; Mismatches 12; Indels 0; Gaps 0;

QY 3 RSIENSLVAMDFSGOKSRVIENPTALSV 32
DB 35 RSISEGCSSLSKFGLESKVVQNLRINV 64

RESULT 23
Q8TKZ4 PRELIMINARY; PRT; 87 AA.
ID Q8TKZ4
AC Q8TKZ4;
DT 01-JUN-2002 (TRENBLrel. 21, Created)
DT 01-JUN-2002 (TRENBLrel. 21, Last sequence update)
DE 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
DE Predicted protein.
GN MA3249.
OS Methanosarcina acetivorans.
OC Archaea; Euryarchaeota; Methanococci; Methanosarcinales;
OC Methanosarcinaceae; Methanosarcina.
OX NCBI_TaxID=2214;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=C2A / ATCC 35395 / DSM 2834;
RX MEDLINE=21929760; PubMed=11932238;
RA Galagan J.E., Nusbaum C., Roy A., Endrizzi M.G., Macdonald P.,
RA FitzHugh W., Calvo S., Engels R., Smirnov S., Atnoor D., Brown A.,
RA Allen N., Naylor J., Stange-Thomann N., DeArellano K., Johnson R.,
RA Linton L., McEwan P., McKernan K., Talamas J., Tirrell A., Ye W.,
RA Zimmer A., Barber R.D., Cann I., Graham D.E., Grahame D.A., Guss A.M.,
RA Hedderich R., Ingram-Smith C., Kuettnner H.C., Krzycki J.A.,
RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
RA Perry J.G., Jarrell K.F., White O., White R.H., de Macario E.C.,
RA Pritchett M., Sowers K.R., Jing H., Macario A.J.L., Paulsen I.,
RA Metcalf W.W., Birren B.;
RT "The genome of Methanosarcina acetivorans reveals extensive metabolic
RT and physiological diversity.";
RL Genome Res. 12:532-542(2002).
DR EMBL; AE011028; AAM06620.1;
RW Complete proteome.
SQ SEQUENCE 87 AA; 10125 MW; 1A970190E4B5CD2D CRC64;

Query Match 19.3%; Score 41; DB 17; Length 87;
Best Local Similarity 22.7%; Pred. No. 6e+02;
Matches 10; Conservative 13; Mismatches 17; Indels 4; Gaps 1;

QY 4 SISENSLVAMDFSGOKSRV----IENPTALSVAVVEGLAWKK 43
DB 5 SLKRTSLTIIRFSTOESSISOKNLKKKEAKIMTIEDSIELKRE 48

RESULT 24
Q74620

```

```

ID O74670 PRELIMINARY; PRT; 68 AA.
AC Q74630;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE Reverse transcriptase (Fragment).
GN POL.
OS Human immunodeficiency virus type 1.
OC Viruses; Retrovirdae; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=P13;
RX MEDLINE=95074869; PubMed=7983713;
RA Najera I., Holguin A., Quinones-Mateu M.E., Munoz-Fernandez M.A.,
RA Najera R., Lopez-Galindez C., Domingo E.;
RT "Pol gene quasiespecies of human immunodeficiency virus: mutations
RT associated with drug resistance in virus from patients undergoing no
RT drug therapy.";
RL J. Virol. 69:23-31(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=P13;
RX MEDLINE=97000986; PubMed=8844016;
RA Quinones-Mateu M.E., Holguin A., Dopazo J., Najera I., Domingo E.;
RT "Point mutant frequencies in the pol gene of human immunodeficiency
RT virus type 1 are two- to threefold lower than those of env.";
RL AIDS Res. Hum. Retroviruses 12:1117-1128(1996).
DR EMBL; U14846; AAC55753.1; -.
DR InterPro; IPR000477; RVTse.
DR Pfam; PF00078; rvt; 1.
KW RNA-directed DNA polymerase.
FT NON_TER 1
FT NON_TER 68
SQ SEQUENCE 68 AA; 7910 MW; 187F11C0FCE674C8 CRC64;

Query Match 19.1%; Score 40.5; DB 15; Length 68;
Best Local Similarity 33.3%; Pred. No. 5.2e+02;
Matches 11; Conservative 7; Mismatches 12; Indels 3; Gaps 2;

QY 13 MDFSQKSRV-IENP--TEALSVAVEEGLAWRK 42
|: |:: ||| |::: |||
Db 1 MEKEGKISRGIPENYNTVPFAIKKDKSTKWRK 33

RESULT 25
Q74630
ID O74630 PRELIMINARY; PRT; 68 AA.
AC Q74630;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE Reverse transcriptase (Fragment).
GN POL.
OS Human immunodeficiency virus type 1.
OC Viruses; Retrovirdae; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=P18;
RX MEDLINE=95074869; PubMed=7983713;
RA Najera I., Holguin A., Quinones-Mateu M.E., Munoz-Fernandez M.A.,
RA Najera R., Lopez-Galindez C., Domingo E.;
RT "Pol gene quasiespecies of human immunodeficiency virus: mutations
RT associated with drug resistance in virus from patients undergoing no
RT drug therapy.";
RL J. Virol. 69:23-31(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=P18;
RX MEDLINE=97000986; PubMed=8844016;
RA Quinones-Mateu M.E., Holguin A., Dopazo J., Najera I., Domingo E.;
RT "Point mutant frequencies in the pol gene of human immunodeficiency

```

```

RT virus type 1 are two- to threefold lower than those of env.";
RL AIDS Res. Hum. Retroviruses 12:1117-1128(1996).
DR EMBL; U14856; AAC55763.1; -.
DR InterPro; IPR000477; RVTse.
DR Pfam; PF00078; rvt; 1.
KW RNA-directed DNA polymerase.
FT NON_TER 1
FT NON_TER 68
SQ SEQUENCE 68 AA; 7938 MW; 18678049FCE674C8 CRC64;

Query Match 19.1%; Score 40.5; DB 15; Length 68;
Best Local Similarity 33.3%; Pred. No. 5.2e+02;
Matches 11; Conservative 7; Mismatches 12; Indels 3; Gaps 2;

QY 13 MDFSQKSRV-IENP--TEALSVAVEEGLAWRK 42
|: |:: ||| |::: |||
Db 1 MEKEGKISRGIPENYNTVPFAIKKDKSTKWRK 33

Search completed: March 28, 2003, 09:08:02
Job time : 66.2126 secs

```

















983 24 11.3 71 1 YVEA\_VACCC  
 984 24 11.3 72 1 GLNI\_BRAJA  
 985 24 11.3 72 1 HTI\_BLAZI  
 986 24 11.3 72 1 IFI\_RHIME  
 987 24 11.3 72 1 Y738\_SYNY3  
 988 24 11.3 72 1 YD21\_THENA  
 989 24 11.3 73 1 GP60\_BPSPI  
 990 24 11.3 73 1 ULI1\_HSVI7  
 991 24 11.3 74 1 HPIS\_RHOGE  
 992 24 11.3 74 1 SECE\_METJA  
 993 24 11.3 75 1 GVAL\_HALNI  
 994 24 11.3 75 1 HPIS\_RHOPE  
 995 24 11.3 75 1 RL28\_BUCAI  
 996 24 11.3 75 1 TALA\_CARMA  
 997 24 11.3 75 1 Y250\_TREPA  
 998 24 11.3 75 1 YC68\_PINTH  
 999 24 11.3 75 1 YD13\_THENA  
 1000 24 11.3 75 1 YHHV\_SALTY

## ALIGNMENTS

RESULT 1  
 Y205\_METJA  
 ID Y205\_METJA STANDARD; PRT; 62 AA.  
 AC Q60262;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical protein MJECLO5.  
 GN MJECLO5.

OS Methanococcus jannaschii.  
 OC Archaea: Euryarchaeota; Methanococci; Methanococcales;  
 OC Methanocaldococcaceae; Methanocaldococcus.  
 OX NCBI\_TaxID=2190;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RC STRAIN-JAL-1 / DSM 2661 / ATCC 43067;  
 RX MEDLINE=96337999; PubMed=8688087;  
 RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,  
 RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,  
 RA Kerlavage A.R., Dougherty B.A., Tomb J.F., Adams M.D., Reich C.I.,  
 RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,  
 RA Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,  
 RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,  
 RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,  
 RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;  
 RT "Complete genome sequence of the methanogenic archaeon, Methanococcus  
 jannaschii.";  
 RL Science 273:1058-1073(1996).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: L77118; AAC37071.1;  
 DR TIGR: MJECLO5;  
 KW Hypothetical protein; Complete proteome.  
 FT DOMAIN 3 15 ILE-RICH  
 SQ SEQUENCE 62 AA; 7327 MW; 1624EC7275EBAD7 CRC64;

Query Match 21.9%; Score 46.5; DB 1; Length 62;  
 Best Local Similarity 28.6%; Pred. No. 14;  
 Matches 12; Conservative 8; Mismatches 21; Indels 1; Gaps 1;

OY 3 RSISENSIVAMDFSGOKSRVIENTEALSVAVEEGLAWKK 43  
 DB 18 KKVAERFLKLESSQGMQWKEIREARERAKKQLEEGIEWAKK 59

RESULT 2  
 RPON\_METJA STANDARD; PRT; 73 AA.  
 AC Q57649;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE DNA-directed RNA polymerase subunit N (EC 2.7.7.6).  
 GN RPON OR MJ0196.  
 OS Methanococcus jannaschii.  
 OC Archaea: Euryarchaeota; Methanococci; Methanococcales;  
 OC Methanocaldococcaceae; Methanocaldococcus.  
 OX NCBI\_TaxID=2190;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RC STRAIN-JAL-1 / DSM 2661 / ATCC 43067;  
 RX MEDLINE=96337999; PubMed=8688087;  
 RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,  
 RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,  
 RA Kerlavage A.R., Dougherty B.A., Tomb J.F., Adams M.D., Reich C.I.,  
 RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,  
 RA Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,  
 RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,  
 RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,  
 RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;  
 RT "Complete genome sequence of the methanogenic archaeon, Methanococcus  
 jannaschii.";  
 RL Science 273:1058-1073(1996).

CC FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION  
 CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS  
 CC SUBSTRATES.  
 CC CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +  
 CC (RNA)(N)  
 CC SIMILARITY: BELONGS TO THE ARCHAEABACTERIA RPON / EUKARYOTIC RPBL0  
 CC RNA POLYMERASE SUBUNIT FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL: U67475; AAB98176.1;  
 DR HSSP: Q26147; 1EF4.  
 DR TIGR: MJ0196;  
 DR InterPro: IPR000268; RNA\_pol\_N.  
 DR Pfam: PF01194; RNA\_pol\_N; 1.  
 DR PROSITE: PS01112; RNA\_POL\_N\_8KD; 1.  
 KW Transferrase; DNA-directed RNA polymerase; Transcription; Zinc;  
 FT METAL-binding; Complete proteome.  
 FT METAL 7 7 ZINC (BY SIMILARITY).  
 FT METAL 10 10 ZINC (BY SIMILARITY).  
 FT METAL 44 44 ZINC (BY SIMILARITY).  
 FT METAL 45 45 ZINC (BY SIMILARITY).  
 SQ SEQUENCE 73 AA; 8695 MW; E716EA06D65B831 CRC64;

Query Match 21.7%; Score 46; DB 1; Length 73;  
 Best Local Similarity 34.4%; Pred. No. 20;  
 Matches 11; Conservative 7; Mismatches 12; Indels 2; Gaps 1;

OY 1 PMRSTSENSIVAMDFSGOKSRVI--ENPTEAL 30  
 DB 4 PIRCFSCGNVIAEVFEYKERILKGNPKDVL 35

RESULT 3  
 Y567\_METJA STANDARD; PRT; 82 AA.  
 ID Y567\_METJA  
 AC Q57987;

DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical protein MJ0567.  
 GN MJ0567.  
 OS Methanococcus jannaschii.  
 OC Archaea; Euryarchaeota; Methanococci; Methanococcales;  
 OC Methanocaldococcaceae; Methanocaldococcus.  
 OX NCBI\_TaxID=2190;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=JAL-3 / DSM 2661 / ATCC 43067;  
 RX MEDLINE=96337999; PubMed=8688087;  
 RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,  
 RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,  
 RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,  
 RA Overbeek R., Kirkness E.F., Weidman J.F., Merrick J.M., Glodek A.,  
 RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Merrick J.M., Nguyen D.,  
 RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,  
 RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,  
 RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;  
 RA "Complete genome sequence of the methanogenic archaeon, Methanococcus  
 jannaschii.";  
 RT Science 273:1058-1073(1996).  
 RL [1]  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; U67505; AAB98558.1; -  
 DT TIGR; MJ0567; -  
 DR  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 82 AA; 8766 MW; 3F3810EEFC9F81CE CRC64;  
 Query Match 19.8%; Score 42; DB 1; Length 82;  
 Best Local Similarity 32.5%; Pred. No. 77;  
 Matches 13; Conservative 8; Mismatches 11; Indels 8; Gaps 2;  
 QY 10 LVAMDFSGKGVIRNQNGPVIIITKSGNIRIGRLAKM 41  
 DB 28 LVSMGINSKLVIRNQNGPVIIITKSGNIRIGRLAKM 67  
 RESULT 4  
 RADC\_STAAU STANDARD; PRT; 82 AA.  
 AC P31337;  
 DT 01-JUL-1993 (Rel. 26, Created)  
 DT 01-JUL-1993 (Rel. 26, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE DNA repair protein radC homolog (25 kDa protein) (Fragment).  
 GN RADC.  
 OS Staphylococcus aureus.  
 OC Bacteria; Firmicutes; Bacillales; Staphylococcus.  
 OX NCBI\_TaxID=1280;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=RN450;  
 RX Murphy E.;  
 RL Submitted (JAN-1986) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP PARTIAL SEQUENCE FROM N.A.  
 RC STRAIN=RN450;  
 RX MEDLINE=84117462; PubMed=6320000;  
 RA Murphy E., Loerdahl S.;  
 RT "Transposition of Tn554 does not generate a target duplication.";  
 RT Nature 307:292-294(1984).  
 CC -!- FUNCTION: INVOLVED IN DNA REPAIR (BY SIMILARITY).  
 CC -!- SIMILARITY: BELONGS TO THE RADC FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; K02985; AAA26680.1; -  
 DR InterPro; IPR001405; RADC.  
 DR ProDom; PD007415; RADC; 1.  
 DR PROSITE; PS01302; RADC; 1.  
 KW DNA repair.  
 FT NON\_TER  
 SQ SEQUENCE 82 AA; 8920 MW; 65E8BF06E3DEC3A4 CRC64;  
 Query Match 18.9%; Score 40; DB 1; Length 82;  
 Best Local Similarity 40.9%; Pred. No. 1.4e+02;  
 Matches 9; Conservative 3; Mismatches 10; Indels 0; Gaps 0;  
 QY 15 FSGKSRVIENTEALSVAVEE 36  
 DB 1 FKGTUNSSIVHPREIFSIAVRE 22  
 RESULT 5  
 RS22\_ECOLI STANDARD; PRT; 45 AA.  
 AC P28690;  
 DT 01-DEC-1992 (Rel. 24, Created)  
 DT 01-DEC-1992 (Rel. 24, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE 30S ribosomal protein S22 (Stationary-phase-induced ribosome-  
 DE associated protein) (SRA) (Protein D).  
 GN RPSV OR SRA OR B1480 OR Z2230 OR ECS2084.  
 OS Escherichia coli, and  
 OS Escherichia coli O157:H7.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia  
 OX NCBI\_TaxID=562, 83334;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-37.  
 RC STRAIN=K12 / W3110;  
 RX MEDLINE=21189300; PubMed=11292794;  
 RA Izutsu K., Wada C., Komine Y., Sako T., Ueguchi C., Nakura S.,  
 RA Wada A.;  
 RT "Escherichia coli ribosome-associated protein SRA, whose copy number  
 RT increases during stationary phase.";  
 RL J. Bacteriol. 183:2765-2773(2001).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12;  
 RX MEDLINE=90337272; PubMed=2199308;  
 RA Mahajan S.K., Chu C.C., Willis D.K., Templin A., Clark A.J.;  
 RT "Physical analysis of spontaneous and mutagen-induced mutants of  
 RT Escherichia coli K-12 expressing DNA exonuclease VIII activity.";  
 RL Genetics 125:261-273(1990).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K12 / MG1655;  
 RX MEDLINE=97426617; PubMed=9278503;  
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Mau B., Shao Y.;  
 RT "The complete genome sequence of Escherichia coli K-12.";  
 RL Science 277:1453-1474(1997).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=O157:H7 / EDL933 / ATCC 700927;  
 RX MEDLINE=21074935; PubMed=11206551;  
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; AF158101; AAD42616.1; -  
DR Hypothetical protein.  
KW  
SQ SEQUENCE 71 AA; 8143 MW; 5D56546D2FADF0C CRC64;

Query Match 18.4%; Score 39; DB 1; Length 71;  
Best Local Similarity 32.4%; Pred. No. 1.6e+02;  
Matches 11; Conservative 5; Mismatches 16; Indels 2; Gaps 1;

OY 1 PMSISLSLVAMDFSGOKSR--VIENTEALSV 32  
Db 26 PLAKSTSEKTVNATLANNRERFCIENDTFVTY 59  
:::|||||:  
:::|||||:

RESULT 7  
CSPP STRCO STANDARD; PRT; 67 RA.

ID CSPP STRCO STANDARD; PRT; 67 RA.  
DC P48819;  
DT 01-FEB-1996 (Rel. 33, Created)  
DD 01-FEB-1996 (Rel. 33, Last sequence update)  
DE 13-JUN-2002 (Rel. 41, Last annotation update)  
DS Cold shock protein scf.  
OS SCOF OR SC00527 OR SCF11\_07C.  
GN Streptomyces coelicolor.  
OC Bacteria; Actinobacteria; Actinobacteriales; Actinobacteridae;  
OCC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomycetes.  
ORX NCBI\_Taxid=1902;  
[1]  
SEQUENCE FROM N.A.  
RC STRAIN=A3(2);  
RR Av-Gay Y., Ravin S., Aharonowitz Y., Cohen G.;  
RL Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.  
RN [2]

SEQUENCE FROM N.A.  
RC STRAIN=A3(2) / MI45;  
RR MEDLINE=21996410; PubMed=12000953;  
TX Bentley S.D., Chater K.F., Cerdono-Tarraga A.-M., Challis G.L.,  
THOMSON N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,  
HARPER D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,  
CRONIN A., Fraser A., Goble A., Hidalgo J., Hornsbey T., Howarth S.,  
HUANG C.H., Kieser T., Larke L., Murphy L., Oliver K., O'Neil S.,  
RABBINOWITZ E., Rajandream M.A., Rutherford K., Rutter S.,  
SEEGER K., Saunders D., Sharp S., Squares R., Taylor K.,  
WARREN T., Wietzorek A., Woodward J., Barrall B.G., Parkhill J.,  
Hopwood D.A.;  
RC "Complete genome sequence of the model actinomycete Streptomyces  
coelicolor A3(2)."  
RR Nature 417:141-147(2002).  
TX Cytoplasmic (Potential).  
CC - SUBCELLULAR LOCATION:  
CC - INDUCTION: IN RESPONSE TO LOW TEMPERATURE.  
CC - SIMILARITY: BELONGS TO THE COLD-SHOCK DOMAIN (CSD) FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; X92686; CAA63367.1; -  
DR EMBL; AL132662; CAB59584.1; -  
RW HSSP; P32081; ICSP.  
OR InterPro: IPR002059; Cold\_shock.  
PFam: PF00313; CSD.1

DR PRINTS: PRO0050; COLDSHOCK.  
 DR PRODOM: PD000621; Cold\_shock; 1.  
 DR SMART: SM00357; CSP; 1.  
 DR PROSITE: PS00352; COLD\_SHOCK; 1.  
 KW Transcription regulation; DNA-binding; Activator; Complete proteome.  
 FT DOMAIN 4 64 CSD.  
 SQ SEQUENCE 67 AA; 7179 MW; E4FDAD9BB1D92B34 CRC64;  
 Query Match 18.2%; Score 38.5; DB 1; Length 67;  
 Best Local Similarity 39.3%; Pred. No. 1.8e+02;  
 Matches 11; Conservative 2; Mismatches 14; Indels 1; Gaps 1;  
 OY 3 RSISENSLVAMDFS-GQKSRVIENTPEA 29  
 DB 40 RELQEGAVTFDITQGGKQPAENITPA 67  
 RESULT 8  
 NIFH\_NOSSN STANDARD; PRT; 74 AA.  
 AC P42336;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Nitrogenase iron protein (EC 1.18.6.1) (Nitrogenase component II)  
 DE (Nitrogenase Fe protein) (Nitrogenase reductase) (Fragment).  
 GN NIFH.  
 OS Nostoc sp. (strain MUN 8820).  
 OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.  
 OX NCBI\_TaxID=55397;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97086627; PubMed=8932316;  
 RA Hill D.R., Belbin T.J., Thorsteinsson M.V., Bassam D., Brass S.,  
 RA Ernst A., Beger P., Paerl H., Mulligan M.E., Potts M.;  
 RT "Gln (cyanoglobins) is a peripheral membrane protein that is  
 RT restricted to certain Nostoc spp.";  
 RL J. Bacteriol. 178:6587-6598(1996).  
 CC CATALYZED BY THE KEY ENZYMIC REACTIONS IN NITROGEN FIXATION ARE  
 CC IRON PROTEIN AND THE MOLYBDENUM-IRON PROTEIN.  
 CC -1- CATALYTIC ACTIVITY: 8 reduced ferredoxin + 8 H(+) + N(2) + 16 ATP  
 CC -1- COFACTOR: BINDS ONE 4FE-4S CLUSTER PER DIMER.  
 CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE NIFH / BCLH / CHLL FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: L47979; AAB41123.1;  
 DR HSPG; P00459; IFFP6.  
 DR InterPro: IPR000392; NitrogenaseII.  
 DR Pfam: PF00142; fer4\_NifH; 1.  
 DR PROSITE: PS00692; NIFH\_FRXC\_2; PARTIAL.  
 DR PROSITE: PS00746; NIFH\_FRXC\_1; PARTIAL.  
 KW Oxidoreductase; Nitrogen fixation; Iron-sulfur; 4Fe-4S; ATP-binding.  
 FT NE\_BIND 13 20 ATP (POTENTIAL).  
 FT NON\_TER 74 74  
 SQ SEQUENCE 74 AA; 7919 MW; 14B88F560242DCDE CRC64;  
 Query Match 18.2%; Score 38.5; DB 1; Length 74;  
 Best Local Similarity 25.5%; Pred. No. 2e+02;  
 Matches 12; Conservative 7; Mismatches 13; Indels 15; Gaps 1;  
 OY 6 SENSILVAMDFSQKSRVI-----ENPTALSVAVEEG 37  
 DB 23 SONTLAAMEGRIILVGCDDPKADSTRMLHLSKAQTVTLHAAERG 69

RESULT 9  
 CINA\_STRGV STANDARD; PRT; 78 AA.  
 AC P29827;  
 DT 01-APR-1993 (Rel. 25, Created)  
 DT 01-APR-1993 (Rel. 25, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Lantibiotic cinnamycin precursor (Lanthiopeptin) (Lantibiotic RO  
 DE 09-0198).  
 GN CINA OR ROCA.  
 OS Streptovorticillum griseovorticillatum.  
 OC Bacteria; Actinobacteria; Actinobacteriales (class); Actinobacteridae;  
 OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomycetes.  
 OX NCBI\_TaxID=68215;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MAR 164C-M16;  
 RX MEDLINE=91301152; PubMed=2070795;  
 RA Kaletta C., Entian K.-D., Jung G.;  
 RA "Peptide sequence of cinnamycin (RO 09-0198): the first structural  
 RT gene of a duramycin-type lantibiotic.";  
 RL Eur. J. Biochem. 199:411-415(1991).  
 RN [2]  
 RP SEQUENCE OF 60-78.  
 RX MEDLINE=91107436; PubMed=2125590;  
 RA Fredenhagen A., Fendrich G., Markl F., Markl W., Gruner J.,  
 RA Raschdorf F., Peter H.H.;  
 RT "Duramycins B and C, two new lantibiotics containing antibiotics as  
 RT inhibitors of phospholipase A2. Structural revision of duramycin and  
 RT cinnamycin.";  
 RL J. Antibiot. 43:1403-1412(1990).  
 RN [3]  
 RP SEQUENCE OF 60-78.  
 RX MEDLINE=89291558; PubMed=2544544;  
 RA Naruse N., Tennyson O., Tomita K., Konishi M., Miyaki T., Kawaguchi H.,  
 RA Fukase K., Wakamiya T., Shiba T.;  
 RT "Lanthiopeptin, a new peptide antibiotic. Production, isolation and  
 RT properties of lanthiopeptin.";  
 RL J. Antibiot. 42:837-845(1989).  
 CC -1- FUNCTION: CAN ACT AS INHIBITOR OF THE ENZYME PHOSPHOLIPASE A2, AND  
 CC OF THE ANGIOTENSIN-CONVERTING ENZYME. SHOWS INHIBITORY ACTIVITIES  
 CC AGAINST HERPES SIMPLEX VIRUS AND IMMUNOPOTENTIATING ACTIVITIES.  
 CC ITS ANTIMICROBIAL ACTIVITIES ARE NOT VERY PRONOUNCED.  
 CC -1- MATURATION OF LANTIBIOTICS INVOLVES THE ENZYMIC CONVERSION OF  
 CC THR, SER, AND CYS INTO DEHYDRATED AA AND THE FORMATION OF SULFIDE  
 CC BRIDGES. THIS IS FOLLOWED BY THE MEMBRANE TRANSLOCATION AND  
 CC CLEAVAGE OF THE MODIFIED PRECURSOR.  
 CC -1- SIMILARITY: STRUCTURAL SIMILARITY TO OTHER TYPE B LANTIBIOTICS.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: X58545; CA941436.1;  
 DR PIR: S17181; EWSMIG.  
 DR PIR: A45767; EWSMCN.  
 KW Antibiotic; Bacteriocin; Lantibiotic.  
 FT PROPEP 1 59 POTENTIAL.  
 FT CHAIN 60 78 LANTIBIOTIC CINNAMYCIN.  
 FT MOD\_RES 63 63 D-ALANINE.  
 FT MOD\_RES 65 65 D-ALANINE (BONDED TO 78).  
 FT MOD\_RES 70 70 D-ABU (AMINOBUTYRIC ACID).  
 FT MOD\_RES 77 77 D-ABU (AMINOBUTYRIC ACID).  
 FT MOD\_RES 78 78 LYSINOALANINE (BONDED TO 65).  
 FT THIOETH 63 73 ALA-S-CYS (LANTHIONINE).  
 FT THIOETH 64 70 ABU-S-CYS (BETA-METHYLLANTHIONINE).  
 FT THIOETH 65 78 ALA-S-CYS (BETA-METHYLLANTHIONINE).





```

CC RNA POLYMERASE SUBUNIT FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use, by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AL445064; CAC11573.1;
CC HSSP: O26147; IEF4.
CC InterPro: IPR000268; RNA_pol_N.
CC Pfam: PF011194; RNA_POL_N_8KD; 1.
CC PROSITE: PS01112; RNA_POL_N_8KD; 1.
CC Transferase: DNA-directed RNA polymerase; Transcription; Zinc;
CC Metal-binding; Complete proteome.
CC METAL 7
CC METAL 10 7 ZINC (BY SIMILARITY).
CC METAL 10 10 ZINC (BY SIMILARITY).
CC METAL 53 53 ZINC (BY SIMILARITY).
CC METAL 54 54 ZINC (BY SIMILARITY).
CC SEQUENCE 72 AA; 8368 MW; 792AEDA20E5447E2 CRC64;
CC -----
CC Query Match 17.9%; Score 38; DB 1; Length 72;
CC Best Local Similarity 30.6%; Pred. No. 2.3e+02;
CC Matches 11; Conservative 5; Mismatches 20; Indels 0; Gaps 0;
CC -----
CC QY 1 PMRSISENSLVAMDFSGQSRVIENTEALSVAVEE 36
CC I: | | | | | | | | | | | | | | | | | |
CC D 4 PVRFCSCGRVIAADYGRYIKRYNEIKAGRDPSPEE 39
CC -----
CC RESULT 13
CC RPN THEVO
CC ID RPN THEVO STANDARD; PRT; 72 AA.
CC AC Q979KO;
CC DT 15-JUN-2002 (Rel. 41, Created)
CC DT 15-JUN-2002 (Rel. 41, Last sequence update)
CC DE 15-JUN-2002 (Rel. 41, Last annotation update)
CC DE DNA-directed RNA polymerase subunit N (EC 2.7.7.6).
CC GN RPN OR TV1161 OR TVG1188103.
CC OS Archaea; Euryarchaeota; Thermoplasmata; Thermoplasmatales;
CC OC Thermoplasmataceae; Thermoplasmata.
CC OX NCBI_TaxID=50339;
CC RN [1]
CC RP SEQUENCE FROM N.A.
CC RC STRAIN=GSS1 / DSM 4299 / JCM 9571;
CC RX MEDLINE=20570466; PubMed=11121031;
CC RA Kawashima T., Amano N., Koike H., Makino S.-I., Higuchi S.,
CC RA Kawashima-Ohya Y., Watanabe K., Yamazaki M., Kanehori K., Kawamoto T.,
CC RA Nunoshiba T., Yamamoto Y., Aramaki H., Makino K., Suzuki M.;
CC RT "Archaeal adaptation to higher temperatures revealed by genomic
CC sequence of Thermoplasma volcanium".
CC RL Proc. Natl. Acad. Sci. U.S.A. 97:14257-14262(2000).
CC CC -1- FUNCTION: DNA-dependent RNA polymerase catalyzes the transcription
CC of DNA into RNA using the four ribonucleoside triphosphates as
CC substrates.
CC CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +
CC (RNA)(N).
CC CC -1- SIMILARITY: BELONGS TO THE ARCHAEABACTERIA RPN / EUKARYOTIC RPB10
CC RNA POLYMERASE SUBUNIT FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use, by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AP000995; BAB60303.1;
CC InterPro: IPR000268; RNA_pol_N.

```

```

DR Pfam: PF01194; RNA_POL_N; 1.
DR PROSITE: PS01112; RNA_POL_N_8KD; 1.
DR Transferase: DNA-directed RNA polymerase; Transcription; Zinc;
DR Metal-binding; Complete proteome.
DR METAL 7
DR METAL 10 7 ZINC (BY SIMILARITY).
DR METAL 10 10 ZINC (BY SIMILARITY).
DR METAL 53 53 ZINC (BY SIMILARITY).
DR METAL 54 54 ZINC (BY SIMILARITY).
DR SEQUENCE 72 AA; 8483 MW; 06AEC0AA7AC75CA6 CRC64;
DR -----
DR Query Match 17.9%; Score 38; DB 1; Length 72;
DR Best Local Similarity 27.8%; Pred. No. 2.3e+02;
DR Matches 10; Conservative 6; Mismatches 20; Indels 0; Gaps 0;
DR -----
DR QY 1 PMRSISENSLVAMDFSGQSRVIENTEALSVAVEE 36
DR I: | | | | | | | | | | | | | | | | | |
DR D 4 PVRFCSCGRVIAADYGRYIKRYNEIRSEGREPTAE 39
DR -----
DR RESULT 14
DR TMOB PSEME
DR ID TMOB PSEME STANDARD; PRT; 83 AA.
DR AC Q00457;
DR DT 01-NOV-1995 (Rel. 32, Created)
DR DT 01-NOV-1995 (Rel. 32, Last sequence update)
DR DT 15-JUN-2002 (Rel. 41, Last annotation update)
DR DE Toluene-4-monooxygenase system protein B (EC 1.14.13.-).
DR GN TMOB.
DR OS Pseudomonas mendocina.
DR OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
DR OC Pseudomonas.
DR OX NCBI_TaxID=300;
DR RN [1]
DR RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-27.
DR RC STRAIN=KRI;
DR RX MEDLINE=91358306; PubMed=1885512;
DR RA Yen K.-M., Karl M.R., Blatt L.M., Simon M.J., Winter R.B.,
DR RA Fausset P.R., Lu H.S., Harcourt A.A., Chen K.K.;
DR RT "Cloning and characterization of a Pseudomonas mendocina KRI gene
DR RT cluster encoding toluene-4-monooxygenase.".
DR RL J. Bacteriol. 173:5315-5327(1991).
DR CC -1- FUNCTION: HYDROXYLATES TOLUENE TO FORM P-CRESOL.
DR CC -1- COFACTOR: FAD; REQUIRES FE(+2) FOR ACTIVITY.
DR CC -1- PATHWAY: Toluene degradation; first step.
DR CC -1- SUBUNIT: THE MULTICOMPONENT ENZYME TOLUENE-4-MONOOXYGENASE
DR IS FORMED BY THE TMOA, TMOB, TMOC, TMOD, TMOE AND TMOF
DR POLYPEPTIDES.
DR -----
DR This SWISS-PROT entry is copyright. It is produced through a collaboration
DR between the Swiss Institute of Bioinformatics and the EMBL outstation -
DR the European Bioinformatics Institute. There are no restrictions on its
DR use, by non-profit institutions as long as its content is in no way
DR modified and this statement is not removed. Usage by and for commercial
DR entities requires a license agreement (See http://www.isb-sib.ch/announce/
DR or send an email to license@isb-sib.ch).
DR -----
DR EMBL: M63106; AAA26000.1;
DR DR Aromatic hydrocarbons catabolism; Oxidoreductase; Flavoprotein;
DR KW Monooxygenase; FAD; Iron.
DR FT INIT_MET 0
DR FT SEQUENCE 83 AA; 9457 MW; 4729FEF73F266F44 CRC64;
DR -----
DR Query Match 17.9%; Score 38; DB 1; Length 83;
DR Best Local Similarity 58.3%; Pred. No. 2.6e+02;
DR Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
DR -----
DR QY 25 NPTEALSVAVEE 36
DR I: | | | | | | | | | | | | | | | | | |
DR D 72 NPTEVIDVVEE 83
DR -----
DR RESULT 15
DR VG85_BPML5

```

```

ID V085_BPML5 STANDARD; PRT; 88 AA.
AC Q05302;
AT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DE 01-FEB-1994 (Rel. 28, Last annotation update)
DE Gene 85 protein (Gp85)...
GN 85.
OS Mycobacteriophage L5.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae;
OC L5-like viruses
OC NCBI_TaxID=31757;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93211282; PubMed=8459766;
RA Hatfull G.F., Sarkis G.J.;
RT "DNA sequence, structure and gene expression of mycobacteriophage L5: a
  phase system for mycobacterial genetics.";
RL Mol. Microbiol. 7:395-403(1993).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL Outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
CC EMBL: Z18946; CAA79461.1;
CC PIR: S31030; S31030.
CC SEQUENCE 88 AA; 10159 MW; 2927841F7D6D4BC CRC64;
SQ
Query Match 17.5%; Score 37; DB 1; Length 88;
Best Local Similarity 50.0%; Pred. No. 3.8e+02;
Matches 8; Conservative 2; Mismatches 6; Indels 0; Gaps 0;
OY 28 EALSVAVEEGLAWKK 43
DB 61 EALFDITWDGVAAWAKK 76
RESULT 16
RS19_CHLPN STANDARD; PRT; 88 AA.
ID RS19_CHLPN
AC Q927R1; Q9JRT8;
AT 30-MAY-2000 (Rel. 39, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE 30S ribosomal protein S19.
GN RPSS OR RS19 OR CPN0643 OR CP0104.
OS Chlamydia pneumoniae (Chlamydophila pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydophila.
OX NCBI_TaxID=83558;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=CW1029;
RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";
RL Nat. Genet. 21:385-389(1999).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=AR39;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
  pneumoniae AR39.";
RL Nucleic Acids Res. 28:1397-1406(2000).
RN [3]

```

```

RP SEQUENCE FROM N.A.
RX STRAIN=J138;
RX MEDLINE=20330349; PubMed=10871362;
RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
  from Japan and CW1029 from USA.";
RL Nucleic Acids Res. 28:2311-2314(2000).
CC -!- FUNCTION: PROTEIN S19 FORMS A COMPLEX WITH S13 THAT BINDS STRONGLY
  TO THE 16S RIBOSOMAL RNA (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE S19P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL Outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
CC EMBL: A5001647; AAD18782.1;
CC DR EMBL: A5002173; AAF73629.1;
CC DR EMBL: AP002547; BAA98850.1;
CC DR HSSP: P80381; 1QKF.
CC DR TIGR: CP0104;
CC InterPro: IPR002222; Ribosomal_S19.
CC Pfam: PF00203; Ribosomal_S19; 1.
CC ProDom: PD001012; Ribosomal_S19; 1.
CC TRFAMS: TIGR01050; rpss_bact; 1.
CC POSITE: PS00323; RIBOSOMAL_S19; 1.
CC Ribosomal protein; rRNA-binding; Complete proteome.
KW CONFLICT 2 2 S -> G (IN REF. 1).
SQ SEQUENCE 88 AA; 10201 MW; EB1E98E3B4BC2588 CRC64;
Query Match 17.2%; Score 36.5; DB 1; Length 88;
Best Local Similarity 26.2%; Pred. No. 4.5e+02;
Matches 11; Conservative 9; Mismatches 21; Indels 1; Gaps 1;
OY 1 PMRSISENSLVAMDFSGKSRVIENPTEALSVAVEEGLAWRK 42
DB 30 PIKTWSRRSMITPEMIGHTFEV-HNGKKELTVFVSETVGHK 70
RESULT 17
RS19_CHLTP STANDARD; PRT; 88 AA.
ID RS19_CHLTP
AC Q84529; Q9PJL8;
AT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE 30S ribosomal protein S19.
GN RPSS OR RS19 OR CT524 OR TC0811.
OS Chlamydia trachomatis, and
OS Chlamydia muridarum.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=813, 83560;
RN [1]
RP SEQUENCE FROM N.A.
RX SPECIES=C.trachomatis; STRAIN=D/UW-3/Cx;
RX MEDLINE=99000809; PubMed=9784136;
RA Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,
RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
RA Davis R.W.;
RT "Genome sequence of an obligate intracellular pathogen of humans:
  Chlamydia trachomatis.";
RL Science 282:754-759(1998).
RN [2]
RP SEQUENCE FROM N.A.
RX SPECIES=C.muridarum; STRAIN=MoPn / N19g;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,

```





```

Db      7 SYDSFSCDRTGRRNAVPIQGDSEAVSV 35
      | : :: | :|:: | | : :|::|
      | : :: | :|:: | | : :|::|

RESULT 22
GCHI_MUCHA
ID GCHI_MUCHA STANDARD; PRT; 80 AA.
AC P51598;
DF 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE GTP cyclohydrolase I (EC 3.5.4.16) (GTP-CH-1) (Fragment).
OS Mucuna hassjoo.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Mucuna.
OX NCBI_TaxID=40337;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95352066; PubMed=7542887;
RA Maier J., Witter K., Guetlich M., Ziegler I., Werner T., Ninnemann H.;
RT "Homology cloning of GTP-cyclohydrolase I from various unrelated
RT eukaryotes by reverse-transcription polymerase chain reaction using a
RT general set of degenerate primers."
RL Biochem. Biophys. Res. Commun. 212:705-711(1995).
CC -1- CATALYTIC ACTIVITY: GTP + 2 H(2)O = formate + 2-amino-4-hydroxy-6-
CC (erythro-1,2,3-trihydroxypropyl) dihydropteridine triphosphate.
CC -1- ENZYME REGULATION: GTP SHOWS A POSITIVE ALLOSTERIC EFFECT. AND
CC TETRAHYDROBIOTERIN INHIBITS THE ENZYME ACTIVITY (BY SIMILARITY).
CC -1- PATHWAY: Tetrahydrobiopterin biosynthesis; first step.
CC -1- SUBUNIT: HOMOPOLYMER (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE GTP CYCLOHYDROLASE I FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: Z49759; CAA89829.1;
CC HSP: P27511; IGTP.
CC InterPro: IPR001474; GTP_cyclohydrol.
CC Pfam: PF01227; GTP_cyclohydrol; 1.
CC ProDom: PD003330; GTP_cyclohydrol; 1.
CC PROSITE: PS00859; GTP_CYCLOHYDROL_1_1; PARTIAL.
CC PROSITE: PS00860; GTP_CYCLOHYDROL_1_2; 1.
CC Tetrahydrobiopterin biosynthesis; Hydrolase; Allosteric enzyme.
KW NON_TER 1 1 1 BY SIMILARITY.
FT DISULFID 4 75
FT NON_TER 80 80
FT SEQUENCE 80 AA; 8845 MW; F5F6C04D2BBDCC36 CRC64;

Query Match 17.0%; Score 36; DB 1; Length 80;
Best Local Similarity 40.0%; Pred. No. 4.7e+02;
Matches 12; Conservative 2; Mismatches 10; Indels 6; Gaps 1;

QY 5 ISENSLVAMDFSGQKSRVNIENPTEALSVAV 34
Db 44 IQEN-----LSRQIAEAVEQVTGALGVAV 67

RESULT 23
TRBG_ECOLI
ID TRBG_ECOLI STANDARD; PRT; 83 AA.
AC P41072;
DF 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE TrbG protein.
GN TRBG.
OS Escherichia coli.

```

```

OG Plasmid F.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94292442; PubMed=8021201;
RA Doran T.J., Loh S.M., Firth N., Skurray R.A.;
RT "Molecular analysis of the F plasmid tracr region: trav encodes a
RT lipoprotein"; 176:4182-4186(1994).
RL J. Bacteriol. 176:4182-4186(1994).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94359430; PubMed=7915817;
RA Frost L.S., Ippen-Ihler K., Skurray R.A.;
RT "Analysis of the sequence and gene products of the transfer region of
RT the F sex factor"; 58:162-210(1994).
RL Microbiol. Rev. 58:162-210(1994).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / CR63;
RA Shimizu H., Saitoh Y., Suda Y., Uehara K., Sampei G., Mizobuchi K.;
RT "Complete nucleotide sequence of the F plasmid: its implications for
RT organization and diversification of plasmid genomes.";
RT Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
RL -1- FUNCTION: UNKNOWN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U01159; AAC44209.1;
CC EMBL: AP001918; BAA97951.1;
CC EcoGene; EG40096; trbg.
CC Plasmid; Complete proteome.
SQ SEQUENCE 83 AA; 9126 MW; D2B935EB7CA26137 CRC64;

Query Match 17.0%; Score 36; DB 1; Length 83;
Best Local Similarity 41.2%; Pred. No. 4.9e+02;
Matches 7; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 10 LVAMDFSGQKSRVNIENP 26
Db 42 IVAVYKSSRSSTVFENP 58

RESULT 24
GEGU_BOVIN
ID GEGU_BOVIN STANDARD; PRT; 69 AA.
AC P50154;
DF 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Guanine nucleotide-binding protein G(I)/G(S)/G(O) gamma-T2 subunit
DE (G-gamma-C) (G-gamma-8).
GN GNGT2 OR GNG8 OR GNGT8.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Retina.
RX MEDLINE=95238332; PubMed=7721746;
RA Ong O.C., Yamane H.K., Phan K.B., Fong H.K., Bok D., Lee R.H.,
RA Fung B.K.-K.;
RT "Molecular cloning and characterization of the G protein gamma
RT subunit of cone photoreceptors.";

```



No.	Score	Match	Length	DB	ID	Description
1	46.5	21.9	62	2	E64510	hypothetical prote
2	46	21.7	76	2	E64324	DNA-directed RNA p
3	45.5	21.5	73	2	T25763	hypothetical prote
4	45	21.2	60	2	A42960	ferrioxin 2(4Fe-4
5	45	21.2	77	2	H69420	hydrogenase expres
6	45	21.2	82	2	D69087	hydrogenase expres
7	42	19.8	53	2	C82776	hypothetical prote
8	42	19.8	79	2	A86517	hypothetical prote
9	42	19.8	79	2	B72106	hypothetical prote
10	42	19.8	81	2	D81565	hypothetical prote
11	42	19.8	82	1	G64370	conserved hypothet
12	41.5	19.6	87	2	S60284	conserved hypothet
13	40.5	19.1	73	2	A60172	B12b protein - bar
14	40	18.9	66	2	A60172	proteoglycan core
15	40	18.9	67	2	A43602	T-cell-stimulating
16	40	18.9	70	2	E70985	endoglucanase (tru
17	39.5	18.6	61	2	F70985	hypothetical prote
18	39.5	18.6	67	2	F86696	4-oxalocitronate t
19	39.5	18.6	77	2	AF1487	probable transcrip
20	39	18.4	45	1	AD1945	hypothetical prote
21	39	18.4	45	1	C64901	ribosomal protein
22	39	18.4	45	2	D90889	30S ribosomal subu
23	39	18.4	45	2	E85728	30S ribosomal subu
24	39	18.4	62	2	T0565A	hypothetical prote
25	39	18.4	72	2	A03532	hypothetical prote
26	39	18.4	77	2	B53269	hypothetical prote
27	39	18.4	87	2	D82861	hypothetical prote
28	39	18.4	90	2	H72072	conserved hypothet
29	39	18.4	90	2	E86552	CT388 hypothetical
30	38.5	18.2	48	2	T35253	small hypothetical















979 28 13.2 83 2 A71710  
 980 28 13.2 83 2 I48283  
 981 28 13.2 84 1 WBPWB  
 982 28 13.2 84 2 G3202  
 983 28 13.2 84 2 G3294  
 984 28 13.2 84 2 G8069  
 985 28 13.2 84 2 G8069  
 986 28 13.2 84 2 A3321  
 987 28 13.2 85 1 IWBV  
 988 28 13.2 85 2 B2832  
 989 28 13.2 85 2 B2695  
 990 28 13.2 85 2 B87422  
 991 28 13.2 85 2 H83399  
 992 28 13.2 85 2 C83392  
 993 28 13.2 85 2 B90048  
 994 28 13.2 85 2 A2412  
 995 28 13.2 86 1 Q1BP67  
 996 28 13.2 86 2 B37402  
 997 28 13.2 86 2 S21595  
 998 28 13.2 86 2 S74567  
 999 28 13.2 86 2 T17806  
 1000 28 13.2 86 2 H75509  
 205982

## ALIGNMENTS

RESULT 1  
 E64324  
 A:Accession: E64324  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-62 <BUL>  
 A:Cross-references: GB:L77118; NID:gl500644; TIGR:MJECLO5; PIDN:AC37071.1; PID:gl500644  
 A:Map position: ECLFOR3265-3453  
 A:Gene: plasmid  
 A:Start codon: GTG  
 A:Note: this stable 58-kilobase pair plasmid is also designated ECL (large extrachromosomal)

Query Match 21.9%; Score 46.5; DB 2; Length 62;  
 Best Local Similarity 28.6%; Pred. No. 28;  
 Matches 12; Conservative 8; Mismatches 21; Indels 1; Gaps 1;  
 Oy 3 RSTSENSLVAMDFS-GOKSRVIENTPEALSVAVEGLAWRK 43  
 Db 18 KKAERFLKLESSQGMWKEIRERAERAKKOLEGIEWAKK 59

RESULT 2  
 E64324  
 A:Accession: E64324  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-62 <BUL>  
 A:Cross-references: GB:L77118; NID:gl500644; TIGR:MJECLO5; PIDN:AC37071.1; PID:gl500644  
 A:Map position: ECLFOR3265-3453  
 A:Gene: plasmid  
 A:Start codon: GTG  
 A:Note: this stable 58-kilobase pair plasmid is also designated ECL (large extrachromosomal)

A:Accession: E64324  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-76 <BUL>  
 A:Cross-references: GB:U67475; GB:L77117; NID:gl590930; PID:gl590941; TIGR:MJ0196; PI  
 A:Map position: FOR190573-190803  
 A:Start codon: GTG  
 A:Superfamily: DNA-directed RNA polymerase II chain RPB10  
 C:Keywords: nucleotidyltransferase; transcription

Query Match 21.7%; Score 46; DB 2; Length 76;  
 Best Local Similarity 34.4%; Pred. No. 41;  
 Matches 11; Conservative 7; Mismatches 12; Indels 2; Gaps 1;  
 Oy 1 PMSISENSLVAMDFS-GOKSRVIENTPEAL 30  
 Db 7 PIRCFSCGNVIAEVFEYKERILKGENPKDVL 38

RESULT 3  
 T25763  
 A:Accession: T25763  
 A:Status: preliminary; translated from GB/EMBL/DBDJ  
 A:Molecule type: DNA  
 A:Residues: 1-73 <PAU>  
 A:Cross-references: EMBL:U88173; PIDN:AB42266.1; GSPDB:GN00019; CESP:F46F11.4  
 A:Experimental source: strain Bristol N2; clone F46F11  
 A:Gene: CESP:F46F11.4  
 A:Map position: 1  
 A:Introns: 38/2

Query Match 21.5%; Score 45.5; DB 2; Length 73;  
 Best Local Similarity 29.4%; Pred. No. 46;  
 Matches 10; Conservative 7; Mismatches 12; Indels 5; Gaps 1;  
 Oy 14 DFSGOKSRVIENTPEALS-----VAVEGLAWRK 42  
 Db 8 DRLGKVKRKNPSDTIGDLKKLIAAQTGRWEK 41

RESULT 4  
 A42960  
 A:Accession: A42960  
 A:Status: preliminary; translated from NCBI backbone (NCBIN:110322, NCBIP:110324)  
 A:Molecule type: DNA  
 A:Residues: 1-60 <CLE>  
 A:Note: sequence extracted from NCBI backbone (NCBIN:110322, NCBIP:110324)  
 C:Genetics:  
 C:Superfamily: ferridoxin 2[4Fe-4S]; ferridoxin 2[4Fe-4S] homology  
 F:3-59/Domain: ferridoxin 2[4Fe-4S] homology <FER>  
 F:10,13,16,51/Binding site: 4Fe-4S cluster (Cys) (covalent) #status predicted  
 F:20,41,44,47/Binding site: 4Fe-4S cluster (Cys) (covalent) #status predicted

Query Match 21.5%; Score 45.5; DB 2; Length 73;  
 Best Local Similarity 29.4%; Pred. No. 46;  
 Matches 10; Conservative 7; Mismatches 12; Indels 5; Gaps 1;  
 Oy 14 DFSGOKSRVIENTPEALS-----VAVEGLAWRK 42  
 Db 8 DRLGKVKRKNPSDTIGDLKKLIAAQTGRWEK 41





;Status: preliminary; nucleic acid sequence not shown; translation not shown  
 ;Molecule type: DNA  
 ;Residues: 1-82 <BOLD>  
 ;Cross-references: GB:U67505; GB:L77117; NID:g2826297; PID:AA090558.1; PID:g1591273; T  
 ;Genetics;  
 ;Map position: REV504744-504496  
 ;Superfamily: Methanococcus jannaschii conserved hypothetical protein MJ0567

```

Query Match      19.1%; Score 40.5; DB 2; Length 73;
Best Local Similarity 38.7%; Pred. No. 2.2e+02;
Matches 12; Conservative 3; Mismatches 7; Indels 9; Gaps 2;

YY      18 QKSRVIENTEALSVAVEEG-----LAW 40
      | | | | | | | | | | | | | |
bb      44 QNSAIATPP-ENLNAATEDGPHQCDAGWLAW 73
      | | | | | | | | | | | | | |

RESULT 14
43602
:Cell-stimulating antigen - Coccidioides immitis (fragment)
:Species: Coccidioides immitis
:date: 29-Jan-1993 #sequence_revision 29-Jan-1993 #text_change 11-Jan-2000
:Accession: A43602; S16764
:Kirkland, T.N.; Zhu, S.; Kruse, D.; Hsu, L.; Seshan, K.R.; Cole, G.T.
:infect. Immun. 59, 3952-3961, 1991
:title: Coccidioides immitis fractions which are antigenic for immune T lymphocytes.
:Reference number: A43602; MUID:92040063; PMID:18440578
:Accession: A43602
:Molecule type: mRNA

```



A:Reference number: AB1807; MUID:21595285; PMID:11759840  
 A:Accession: AD1945  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-77 <KUR>  
 A:Cross-references: GB:BA000019; PIDN:BAB73068.1; PID:g17130457; GSPDB:GN00179  
 A:Experimental source: strain PCC 7120  
 C:Genetics:  
 A:Gene: asl1111

Query Match 18.6%; Score 39.5; DB 2; Length 77;  
 Best Local Similarity 22.2%; Pred. No. 3.2e+02;  
 Matches 8; Conservative 10; Mismatches 13; Indels 5; Gaps 1;

QY 2 MRSISESLVAMDFSGKSRVIENTEALSVAVEEG 37  
 DB 10 LKAVKQALR-----ERLQATNNPEAFIKAQEEG 40

RESULT 20  
 C64901  
 ribosomal protein S22 [validated] - Escherichia coli (strain K-12)  
 C:Species: Escherichia coli  
 C:Date: 24-Sep-1999 #sequence\_revision 24-Sep-1999 #text\_change 01-Mar-2002  
 C:Accession: C64901  
 R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Cohen, A.; Rose, D.J.; Mau, B.; Shao, Y.  
 Science 277, 1453-1462, 1997  
 A:Title: The complete genome sequence of Escherichia coli K-12.  
 A:Reference number: A64720; MUID:97426617; PMID:9278503  
 A:Accession: C64901  
 A:Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-45 <BLAT>  
 A:Cross-references: GB:AE000245; GB:U00096; NID:g1787752; PIDN:AAC74553.1; PID:g1787755;  
 A:Experimental source: strain K-12, substrain MG1655  
 R:Arnold, R.J.; Belli, J.P.  
 Anal. Biochem. 269, 106-112, 1999  
 A:Title: Observation of Escherichia coli ribosomal proteins and their posttranslational  
 A:Reference number: A59071; MUID:99196679; PMID:10094780  
 A:Contents: annotation; mass spectrographic analysis  
 A:Note: a ribosomal protein with these mass spectrographic characteristics was found; no  
 may have been missed  
 C:Genetics:  
 A:Gene: rpsv  
 C:Complex: the ribosome is composed of the large (50S) and small (30S) subunit; the large  
 S rRNA and 22 distinct proteins  
 C:Complex: small subunit ribosomal proteins: S1 (PIR:R3EC1), S2 (PIR:R3EC2), S3 (PIR:R3EC10), S11 (PIR:R3EC11), S12 (PIR:R3EC12), S13 (PIR:R3EC13), S14 (PIR:R3EC14), S15 (PIR:R3EC21), S22 (PIR:C64901) [validated, MUID:99196679]  
 C:Function:  
 A:Pathway: protein biosynthesis  
 C:Superfamily: Escherichia coli ribosomal protein S22  
 C:Keywords: protein biosynthesis; ribosome  
 F:1-45/Product: ribosomal protein S22 #status experimental <MAT>

Query Match 18.4%; Score 39; DB 1; Length 45;  
 Best Local Similarity 63.6%; Pred. No. 2e+02;  
 Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 17 GOKSRVIENTPT 27  
 DB 27 GOKSSVVNNPT 37

RESULT 21  
 D90889  
 30S ribosomal subunit protein S22 [imported] - Escherichia coli (strain O157:H7, substrain  
 C:Species: Escherichia coli  
 C:Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 17-May-2002  
 C:Accession: D90889  
 R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;  
 gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.

DNA Res. 8, 11-22, 2001  
 A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and 9  
 A:Reference number: A99629; MUID:21156231; PMID:11258796  
 A:Accession: D90889  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-45 <HAY>  
 A:Cross-references: GB:BA000007; PIDN:BA035507.1; PID:g13361550; GSPDB:GN00154  
 A:Experimental source: strain O157:H7, substrain RMD 0509952  
 C:Genetics:  
 A:Gene: ECs2084  
 C:Superfamily: Escherichia coli ribosomal protein S22

Query Match 18.4%; Score 39; DB 2; Length 45;  
 Best Local Similarity 63.6%; Pred. No. 2e+02;  
 Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 17 GOKSRVIENTPT 27  
 DB 27 GOKSSVVNNPT 37

RESULT 22  
 E85728  
 30S ribosomal subunit protein S22 [imported] - Escherichia coli (strain O157:H7, subs  
 C:Species: Escherichia coli  
 C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 17-May-2002  
 C:Accession: E85728  
 R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May  
 iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apoda  
 Nature 409, 529-533, 2001  
 A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
 A:Reference number: A83480; MUID:21074935; PMID:11206551  
 A:Accession: E85728  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-45 <STO>  
 A:Cross-references: GB:AE005174; NID:g12515201; PIDN:AAG56289.1; GSPDB:GN00145; UMG  
 A:Experimental source: strain O157:H7, substrain EBL933  
 C:Genetics:  
 A:Gene: rpsv  
 C:Superfamily: Escherichia coli ribosomal protein S22

Query Match 18.4%; Score 39; DB 2; Length 45;  
 Best Local Similarity 63.6%; Pred. No. 2e+02;  
 Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 17 GOKSRVIENTPT 27  
 DB 27 GOKSSVVNNPT 37

RESULT 23  
 T06654  
 hypothetical protein T6G15.70 - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 22-Oct-1999  
 C:Accession: T06654  
 R:Bevan, M.; Murphy, G.; Ridley, P.; Hudson, S.; Bancroft, I.; Mewes, H.W.; Mayer, K.  
 submitted to the Protein Sequence Database, April 1999  
 A:Reference number: 215791  
 A:Accession: T06654  
 A:Molecule type: DNA  
 A:Residues: 1-62 <BEV>  
 A:Cross-references: EMBL:AL049656; GSPDB:GN00062; ATSP:T6G15.70  
 A:Experimental source: cultivar Columbia; BAC clone T6G15  
 C:Genetics:  
 A:Gene: ATSP:T6G15.70  
 A:Map position: 4

Query Match 18.4%; Score 39; DB 2; Length 62;  
 Best Local Similarity 29.7%; Pred. No. 2.9e+02;  
 Matches 11; Conservative 7; Mismatches 19; Indels 0; Gaps 0;

QY 2 MRSISESLVAMDFSGKSRVIENTEALSVAVEEGL 38  
 || : : | || : : : : || : : ||  
 Db 1 MRPWOLDMLSEMDAGSSMAMVDVDDLEAMEILNEGGL 37

RESULT 24

AD3532  
 Hypothetical protein BMEI10182 [imported] - Brucella melitensis (strain 16M)  
 C:Species: Brucella melitensis  
 C>Date: 01-Feb-2002 #sequence\_revision 01-Feb-2002 #text\_change 01-Feb-2002  
 C:Accession: AD3532  
 R:DelVecchio, V.G.; Kaprat, R.J.; Redkar, G.; Patra, C.; Ios, T.; Ivanova,  
 ; Mazur, M.; Gotsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letess,  
 Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002  
 A:Title: The genome sequence of the facultative intracellular pathogen Brucella melitensis  
 A:Reference number: AD3532; PMID:11756688  
 A:Accession: AD3532  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-72 <KUR>  
 A:Cross-references: GB:AE008918; PIDN:AAL53423.1; PID:gl7984319; GSPDB:GN00191  
 A:Experimental source: strain 16M  
 C:Genetics:  
 A:Gene: BMEI10182  
 A:Map position: II

Query Match 18.4%; Score 39; DB 2; Length 72;  
 Best Local Similarity 41.7%; Pred. No. 3.5e+02;  
 Matches 10; Conservative 2; Mismatches 12; Indels 0; Gaps 0;

QY 17 GQSRVIENTEALSVAVEEGLAW 40  
 | | | | | : : || ||  
 Db 4 GHLSYVRRLVESRRRLMVEIGLKW 27

RESULT 25

B83269  
 Hypothetical protein PA3009 [imported] - Pseudomonas aeruginosa (strain PA01)  
 C:Species: Pseudomonas aeruginosa  
 C>Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
 C:Accession: B83269  
 R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Br  
 adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,  
 ; Lory, S.; Olson, M.V.  
 Nature 406, 959-964, 2000  
 A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho  
 A:Reference number: AB2950; MUID:20437337; PMID:10984043  
 A:Accession: B83269  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-77 <STO>  
 A:Cross-references: GB:AE004726; GB:AE004091; NID:g9949108; PIDN:AAG06397.1; GSPDB:GN001  
 A:Experimental source: strain PA01  
 C:Genetics:  
 A:Gene: PA3009

Query Match 18.4%; Score 39; DB 2; Length 77;  
 Best Local Similarity 31.8%; Pred. No. 3.7e+02;  
 Matches 7; Conservative 6; Mismatches 9; Indels 0; Gaps 0;

QY 3 RSISESLVAMDFSGKSRVIE 24  
 || : : | || : : : : || : : ||  
 Db 20 RADDEALVLEFGDAKNFLQ 41

Search completed: March 28, 2003, 09:08:54  
 Job time : 42.2677 secs

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 28, 2003, 09:08:20 ; Search time 13.8819 Seconds  
(without alignments)  
181.900 Million cell updates/sec

File: US09-936-697-5

Sequence: 1 PMRISISLVAMDFSGQK.....ENPTALSAVEGLAWRK 43

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 237916 seqs, 58723674 residues

Total number of hits satisfying chosen parameters: 29382

Minimum DB seq length: 40

Maximum DB seq length: 90

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 1000 summaries

Database: Published Applications: AA

1: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pdb: \*  
2: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pdb: \*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_PUB.pdb: \*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUB.pdb: \*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pdb: \*  
6: /cgn2\_6/ptodata/1/pubpaa/US07\_PUB.pdb: \*  
7: /cgn2\_6/ptodata/1/pubpaa/US07\_PUB.pdb: \*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUB.pdb: \*  
9: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pdb: \*  
10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUB.pdb: \*  
11: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pdb: \*  
12: /cgn2\_6/ptodata/1/pubpaa/US10\_PUB.pdb: \*  
13: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pdb: \*  
14: /cgn2\_6/ptodata/1/pubpaa/US60\_PUB.pdb: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	45	21.2	84	10	US-09-864-761-47521
2	43.5	20.5	55	10	US-09-879-957-111
3	42	19.8	63	10	US-09-864-761-34262
4	42	19.8	72	9	US-09-738-626-6764
5	40	18.9	49	10	US-09-864-761-41352
6	40	18.9	63	9	US-09-864-761-41352
7	40	18.9	70	9	US-09-864-761-41352
8	40	18.9	71	9	US-09-864-761-41352
9	39.5	18.6	68	10	US-09-864-761-47104
10	39.5	18.6	78	10	US-09-764-897-218
11	39.5	18.6	89	10	US-09-815-242-4918
12	39	18.4	47	10	US-09-864-761-43815
13	38.5	18.2	66	10	US-09-864-761-43291
14	38.5	18.2	66	10	US-09-864-761-36073
15	38.5	18.2	85	9	US-09-522-334-28
16	38	17.9	62	10	US-09-939-980-301
17	38	17.9	68	10	US-09-925-300-1560
18	37.5	17.7	51	10	US-09-864-761-43411
19	37.5	17.7	64	10	US-09-925-300-1765

20	37.5	17.7	65	9	US-10-092-154-735
21	37.5	17.7	65	10	US-09-764-847-735
22	37.5	17.7	89	10	US-09-925-297-543
23	37	17.5	57	10	US-09-071-838-294
24	37	17.5	56	10	US-09-764-877-1480
25	36.5	17.2	59	10	US-09-867-550-980
26	36.5	17.2	61	10	US-09-879-957-208
27	36.5	17.2	71	9	US-10-001-876-151
28	36.5	17.2	79	10	US-09-864-761-38542
29	36.5	17.2	83	9	US-09-908-931B-50
30	36.5	17.2	90	9	US-09-764-872-374
31	36	17.0	46	10	US-09-864-761-33369
32	36	17.0	57	10	US-09-864-761-40621
33	36	17.0	58	10	US-09-864-761-47122
34	36	17.0	63	9	US-10-083-357-900
35	36	17.0	64	10	US-09-764-869-864
36	36	17.0	74	9	US-09-738-626-4797
37	36	17.0	78	10	US-09-864-761-33623
38	36	17.0	89	10	US-09-864-761-46485
39	35.5	16.7	45	10	US-09-864-761-43127
40	35.5	16.7	56	9	US-09-742-096-24
41	35.5	16.7	56	9	US-09-742-096-24
42	35.5	16.7	76	10	US-09-864-761-37234
43	35.5	16.7	85	10	US-09-731-872-476
44	35.5	16.7	89	9	US-10-102-806-769
45	35	16.5	43	9	US-10-116-252-26
46	35	16.5	58	10	US-09-925-302-651
47	35	16.5	59	10	US-09-864-761-39483
48	35	16.5	62	10	US-09-764-877-1136
49	35	16.5	64	10	US-09-864-761-48795
50	35	16.5	65	10	US-09-867-550-1238
51	35	16.5	75	10	US-09-864-761-44991
52	35	16.5	76	9	US-10-092-154-881
53	35	16.5	76	10	US-09-764-847-881
54	35	16.5	86	10	US-09-864-761-35345
55	35	16.5	87	10	US-09-864-761-43703
56	34.5	16.3	58	10	US-09-982-610-30
57	34.5	16.3	60	9	US-09-774-639-237
58	34.5	16.3	60	9	US-09-969-730-305
59	34.5	16.3	74	9	US-09-969-730-155
60	34.5	16.3	75	9	US-09-774-639-125
61	34.5	16.3	78	10	US-09-864-761-36451
62	34.5	16.3	83	10	US-09-864-761-41234
63	34	16.0	55	10	US-09-864-761-45639
64	34	16.0	63	10	US-09-864-761-46037
65	34	16.0	65	10	US-09-864-761-36194
66	34	16.0	65	10	US-09-864-761-37813
67	34	16.0	69	10	US-09-864-761-47955
68	34	16.0	75	10	US-09-864-761-36328
69	34	16.0	77	9	US-10-033-066A-75
70	34	16.0	80	10	US-09-867-550-1756
71	34	16.0	90	10	US-09-864-761-41728
72	33.5	15.8	50	10	US-09-864-761-41057
73	33.5	15.8	51	10	US-09-925-297-508
74	33.5	15.8	57	10	US-09-815-242-12724
75	33.5	15.8	65	10	US-09-864-761-43927
76	33.5	15.8	74	9	US-09-796-692-1469
77	33.5	15.8	74	9	US-09-796-692-1911
78	33.5	15.8	74	9	US-09-796-692-1911
79	33.5	15.8	80	10	US-09-864-761-48011
80	33	15.6	40	10	US-09-864-761-34563
81	33	15.6	42	10	US-09-864-761-39032
82	33	15.6	42	10	US-09-864-761-35168
83	33	15.6	55	9	US-09-158-722-2
84	33	15.6	55	10	US-09-864-761-40879
85	33	15.6	62	10	US-09-864-761-48027
86	33	15.6	64	10	US-09-864-761-34561
87	33	15.6	64	10	US-09-864-761-34993
88	33	15.6	69	9	US-09-925-299-1037
89	33	15.6	65	10	US-09-864-761-39839
90	33	15.6	65	10	US-09-925-299-1037
91	33	15.6	67	10	US-09-864-761-38332
92	33	15.6	67	10	US-09-820-893-109

Sequence 735, App  
Sequence 735, App  
Sequence 543, App  
Sequence 294, App  
Sequence 1480, App  
Sequence 980, App  
Sequence 208, App  
Sequence 151, App  
Sequence 38452, A  
Sequence 50, Appl  
Sequence 374, App  
Sequence 33369, A  
Sequence 40621, A  
Sequence 47122, A  
Sequence 900, App  
Sequence 864, App  
Sequence 4797, App  
Sequence 33623, A  
Sequence 46485, A  
Sequence 43127, A  
Sequence 21, Appl  
Sequence 24, Appl  
Sequence 37234, A  
Sequence 476, App  
Sequence 769, App  
Sequence 26, Appl  
Sequence 651, App  
Sequence 39483, A  
Sequence 1136, App  
Sequence 48795, A  
Sequence 1238, App  
Sequence 44991, A  
Sequence 881, App  
Sequence 881, App  
Sequence 35345, A  
Sequence 43703, A  
Sequence 30, Appl  
Sequence 237, App  
Sequence 305, App  
Sequence 155, App  
Sequence 125, App  
Sequence 16451, A  
Sequence 41234, A  
Sequence 45639, A  
Sequence 46037, A  
Sequence 36194, A  
Sequence 37813, A  
Sequence 47955, A  
Sequence 36328, A  
Sequence 75, Appl  
Sequence 1756, App  
Sequence 41728, A  
Sequence 41057, A  
Sequence 508, App  
Sequence 12724, A  
Sequence 43927, A  
Sequence 1469, App  
Sequence 1911, App  
Sequence 1951, App  
Sequence 48011, A  
Sequence 34563, A  
Sequence 39032, A  
Sequence 35168, A  
Sequence 2, Appl  
Sequence 40879, A  
Sequence 48027, A  
Sequence 34561, A  
Sequence 34993, A  
Sequence 1037, App  
Sequence 39839, A  
Sequence 1037, App  
Sequence 38332, A  
Sequence 109, App

93	33	15.6	67	10	US-09-873-880-20	Sequence 20, Appl	166	32	89	10	US-09-867-550-212	Sequence 212, App	
94	33	15.6	68	10	US-09-795-501-8	Sequence 8, Appl	167	32	90	10	US-09-864-761-40114	Sequence 40114, A	
95	33	15.6	69	9	US-10-083-357-694	Sequence 694, App	168	31.5	54	10	US-09-864-761-46492	Sequence 46492, A	
96	33	15.6	70	10	US-09-864-761-37542	Sequence 37542, A	169	31.5	55	9	US-10-078-090-129	Sequence 129, App	
97	33	15.6	71	12	US-10-014-774-9	Sequence 9, Appl	170	31.5	61	10	US-09-864-761-38193	Sequence 38193, A	
98	33	15.6	72	10	US-09-864-761-38891	Sequence 38891, A	171	31.5	61	10	US-09-864-761-48173	Sequence 48173, A	
99	33	15.6	73	10	US-09-864-761-44101	Sequence 44101, A	172	31.5	65	10	US-09-864-761-39828	Sequence 39828, A	
100	33	15.6	74	10	US-09-898-416-17	Sequence 17, Appl	173	31.5	75	10	US-09-921-397-104	Sequence 104, App	
101	33	15.6	78	9	US-10-002-344-206	Sequence 206, App	174	31.5	77	10	US-09-867-550-1256	Sequence 1256, App	
102	33	15.6	78	9	US-10-002-344-207	Sequence 207, App	175	31	40	9	US-10-091-572-223	Sequence 223, App	
103	33	15.6	81	10	US-09-867-550-210	Sequence 210, App	176	31	43	9	US-10-001-835-191	Sequence 191, App	
104	33	15.6	84	9	US-10-013-379-18	Sequence 18, Appl	177	31	44	12	US-10-042-417-72	Sequence 417, Appl	
105	33	15.6	86	10	US-09-864-761-36719	Sequence 36719, A	178	31	44	9	US-10-114-893-62	Sequence 62, Appl	
106	33	15.6	87	9	US-10-083-357-921	Sequence 921, App	179	31	46	46	10	US-09-864-761-42310	Sequence 42310, A
107	32.5	15.3	42	10	US-09-764-869-888	Sequence 888, App	180	31	54	9	US-09-158-722-51	Sequence 51, Appl	
108	32.5	15.3	49	10	US-09-864-761-42434	Sequence 42434, A	181	31	54	10	US-09-864-761-46487	Sequence 46487, A	
109	32.5	15.3	50	10	US-09-071-836-167	Sequence 167, App	182	31	55	9	US-09-158-722-8	Sequence 8, Appl	
110	32.5	15.3	53	10	US-09-764-898-273	Sequence 273, App	183	31	55	10	US-09-864-761-33703	Sequence 33703, A	
111	32.5	15.3	61	10	US-09-864-761-35421	Sequence 35421, A	184	31	55	10	US-09-815-626-12	Sequence 12, Appl	
112	32.5	15.3	63	10	US-09-864-761-44168	Sequence 44168, A	185	31	57	9	US-09-736-692-1598	Sequence 1598, App	
113	32.5	15.3	64	10	US-09-925-300-1332	Sequence 1332, App	186	31	58	10	US-09-864-761-35718	Sequence 35718, App	
114	32.5	15.3	64	10	US-09-864-761-44168	Sequence 44168, A	187	31	58	10	US-09-864-761-38813	Sequence 38813, A	
115	32.5	15.3	71	10	US-09-829-482-10	Sequence 10, Appl	188	31	62	8	US-08-808-031A-19	Sequence 19, Appl	
116	32.5	15.3	72	10	US-09-864-761-35035	Sequence 35035, A	189	31	62	10	US-09-864-761-33340	Sequence 33340, A	
117	32.5	15.3	72	10	US-09-929-924-17	Sequence 17, Appl	190	31	62	10	US-09-864-761-45366	Sequence 45366, A	
118	32.5	15.3	74	9	US-10-092-154-806	Sequence 806, App	191	31	65	9	US-09-864-761-47557	Sequence 47557, A	
119	32.5	15.3	77	10	US-09-764-847-806	Sequence 806, App	192	31	65	9	US-09-966-262-241	Sequence 241, App	
120	32.5	15.3	77	10	US-09-864-761-38805	Sequence 38805, A	193	31	65	10	US-09-864-761-40775	Sequence 40775, A	
121	32.5	15.3	83	10	US-09-908-711-103	Sequence 103, App	194	31	68	10	US-09-764-887-272	Sequence 272, App	
122	32.5	15.3	83	10	US-09-864-761-41142	Sequence 41142, A	195	31	70	10	US-09-765-272-142	Sequence 142, App	
123	32.5	15.3	86	12	US-09-764-877-1675	Sequence 1675, App	196	31	70	10	US-09-867-550-1634	Sequence 1634, App	
124	32.5	15.3	86	12	US-10-153-064-69	Sequence 69, Appl	197	31	77	9	US-10-112-793-26	Sequence 26, Appl	
125	32.5	15.3	88	10	US-09-864-761-48010	Sequence 48010, A	198	31	77	10	US-09-867-550-114	Sequence 114, App	
126	32	15.1	51	10	US-09-864-761-45355	Sequence 45355, A	199	31	78	9	US-09-738-628-5645	Sequence 5645, App	
127	32	15.1	51	10	US-09-864-761-47302	Sequence 47302, A	200	31	78	9	US-10-078-770-184	Sequence 184, App	
128	32	15.1	52	9	US-09-766-692-1579	Sequence 1579, App	201	31	78	9	US-09-815-232-1158	Sequence 1158, A	
129	32	15.1	53	9	US-10-001-863-126	Sequence 126, App	202	31	78	10	US-09-864-761-36144	Sequence 36144, A	
130	32	15.1	54	9	US-09-920-280	Sequence 280, App	203	31	81	10	US-09-764-877-1409	Sequence 1409, App	
131	32	15.1	54	9	US-10-092-154-862	Sequence 862, App	204	31	85	9	US-10-072-783-5	Sequence 5, Appl	
132	32	15.1	54	10	US-09-864-761-46759	Sequence 46759, A	205	31	86	9	US-10-156-275-34	Sequence 34, Appl	
133	32	15.1	54	10	US-09-764-847-862	Sequence 847, App	206	31	86	10	US-09-751-100B-48	Sequence 48, Appl	
134	32	15.1	55	10	US-09-864-761-46169	Sequence 46169, A	207	31	90	10	US-09-922-261-304	Sequence 304, App	
135	32	15.1	55	10	US-09-864-761-46169	Sequence 46169, A	208	31	45	10	US-09-205-658-69	Sequence 69, Appl	
136	32	15.1	57	10	US-09-864-761-44565	Sequence 44565, A	209	30.5	45	10	US-09-844-353A-69	Sequence 69, Appl	
137	32	15.1	57	10	US-09-864-761-48529	Sequence 48529, A	210	30.5	48	10	US-09-864-761-46381	Sequence 46381, A	
138	32	15.1	58	9	US-09-983-802-549	Sequence 549, App	211	30.5	50	9	US-10-083-357-921	Sequence 921, App	
139	32	15.1	58	10	US-09-879-957-131	Sequence 131, App	212	30.5	57	9	US-10-012-896-561	Sequence 561, App	
140	32	15.1	60	10	US-09-815-242-5488	Sequence 5488, App	213	30.5	57	9	US-09-895-783-561	Sequence 561, App	
141	32	15.1	60	10	US-09-815-242-12554	Sequence 12554, A	214	30.5	57	10	US-09-759-143-561	Sequence 561, App	
142	32	15.1	62	9	US-09-796-692-1724	Sequence 1724, App	215	30.5	57	10	US-09-822-827-561	Sequence 561, App	
143	32	15.1	62	9	US-09-309-196-117	Sequence 117, App	216	30.5	57	10	US-09-832-827-561	Sequence 561, App	
144	32	15.1	62	10	US-09-864-761-38019	Sequence 38019, A	217	30.5	57	10	US-09-780-669-561	Sequence 561, App	
145	32	15.1	62	10	US-09-820-893-91	Sequence 91, Appl	218	30.5	60	10	US-09-864-761-37466	Sequence 37466, A	
146	32	15.1	63	10	US-09-820-893-79	Sequence 79, Appl	219	30.5	62	10	US-09-864-761-42598	Sequence 42598, A	
147	32	15.1	65	12	US-10-078-929-24	Sequence 24, Appl	220	30.5	63	10	US-09-864-761-40078	Sequence 40078, A	
148	32	15.1	68	10	US-09-925-299-1186	Sequence 1186, App	221	30.5	63	10	US-09-864-761-41505	Sequence 41505, A	
149	32	15.1	69	10	US-09-925-299-1186	Sequence 1186, App	222	30.5	64	10	US-09-925-301-1339	Sequence 1339, App	
150	32	15.1	70	10	US-09-220-920-98	Sequence 98, Appl	223	30.5	64	10	US-09-867-550-1098	Sequence 1098, App	
151	32	15.1	70	10	US-09-820-893-52	Sequence 52, Appl	224	30.5	65	10	US-09-785-272-146	Sequence 146, App	
152	32	15.1	71	10	US-09-864-761-44183	Sequence 44183, A	225	30.5	68	10	US-09-864-761-43485	Sequence 43485, A	
153	32	15.1	71	10	US-09-864-761-44403	Sequence 44403, A	226	30.5	69	10	US-09-798-652-2308	Sequence 2308, App	
154	32	15.1	72	10	US-09-810-560-7	Sequence 7, Appl	227	30.5	70	10	US-09-881-752A-332	Sequence 332, App	
155	32	15.1	74	9	US-10-001-887-135	Sequence 135, App	228	30.5	70	12	US-10-062-023-3	Sequence 3, Appl	
156	32	15.1	74	10	US-09-864-761-34188	Sequence 34188, A	229	30.5	72	10	US-09-783-320-10	Sequence 10, Appl	
157	32	15.1	76	9	US-09-984-245-239	Sequence 239, App	230	30.5	72	10	US-09-848-664-22	Sequence 22, Appl	
158	32	15.1	76	9	US-10-066-127-6	Sequence 6, Appl	231	30.5	76	10	US-09-963-896-5	Sequence 5, Appl	
159	32	15.1	76	9	US-09-966-262-239	Sequence 262, App	232	30.5	77	9	US-10-072-349-141	Sequence 141, App	
160	32	15.1	79	10	US-09-864-761-42162	Sequence 42162, A	233	30.5	77	10	US-09-764-855-141	Sequence 141, App	
161	32	15.1	81	9	US-10-002-344A-213	Sequence 213, App	234	30.5	82	10	US-09-864-761-42490	Sequence 42490, A	
162	32	15.1	81	10	US-09-410-194-10	Sequence 10, Appl	235	30.5	87	10	US-09-867-550-1468	Sequence 1468, App	
163	32	15.1	83	10	US-09-867-550-1736	Sequence 1736, App	236	30.5	89	9	US-09-309-196-1	Sequence 1, Appl	
164	32	15.1	83	10	US-09-764-877-1070	Sequence 1070, App	237	30.5					
165	32	15.1	84	10	US-09-764-877-1985	Sequence 1985, App	238	30.5					

239	30.5	14.4	89	10	US-09-925-300-964	Sequence 964, App	312	29.5	13.9	50	10	US-09-864-761-37153	Sequence 37153, A
240	30	14.2	41	10	US-09-887-586A-34	Sequence 34, Appl	313	29.5	13.9	50	10	US-09-864-761-46892	Sequence 46892, A
241	30	14.2	41	10	US-09-903-012-34	Sequence 34, Appl	314	29.5	13.9	53	10	US-09-995-494-112	Sequence 112, A
242	30	14.2	42	10	US-09-843-845-30	Sequence 30, Appl	315	29.5	13.9	55	10	US-09-864-761-45582	Sequence 45582, A
243	30	14.2	42	10	US-09-864-761-33922	Sequence 33922, A	316	29.5	13.9	60	10	US-09-864-761-43942	Sequence 43942, A
244	30	14.2	44	9	US-10-153-159-10	Sequence 10, Appl	317	29.5	13.9	61	10	US-09-921-397-33	Sequence 33, Appl
245	30	14.2	44	9	US-10-153-176-10	Sequence 10, Appl	318	29.5	13.9	63	9	US-09-738-626-3628	Sequence 3628, Ap
246	30	14.2	44	9	US-09-921-144-2	Sequence 2, Appl	319	29.5	13.9	66	9	US-09-796-692-2165	Sequence 2165, Ap
247	30	14.2	46	9	US-09-989-919-115	Sequence 115, App	320	29.5	13.9	66	9	US-10-091-483-224	Sequence 224, App
248	30	14.2	46	9	US-10-091-572-327	Sequence 327, App	321	29.5	13.9	66	12	US-10-074-846-224	Sequence 39, Appl
249	30	14.2	46	9	US-09-779-451-42	Sequence 42, Appl	322	29.5	13.9	66	12	US-10-028-415-39	Sequence 33, Appl
250	30	14.2	47	9	US-10-000-356A-167	Sequence 167, App	323	29.5	13.9	66	10	US-09-864-761-48084	Sequence 48084, A
251	30	14.2	49	10	US-09-864-761-34428	Sequence 34428, A	324	29.5	13.9	70	10	US-09-864-761-38015	Sequence 38015, A
252	30	14.2	49	10	US-09-864-761-48537	Sequence 48537, A	325	29.5	13.9	72	10	US-09-864-761-1500	Sequence 1500, Ap
253	30	14.2	50	10	US-09-864-761-34120	Sequence 34120, A	326	29.5	13.9	72	10	US-09-764-877-1500	Sequence 43814
254	30	14.2	50	10	US-09-864-761-41577	Sequence 41577, A	327	29.5	13.9	76	9	US-09-864-761-43814	Sequence 43814, A
255	30	14.2	51	10	US-09-864-761-36508	Sequence 36508, A	328	29.5	13.9	76	9	US-10-178-213-95	Sequence 92, Appl
256	30	14.2	51	10	US-09-864-761-46322	Sequence 46322, A	329	29.5	13.9	76	9	US-10-178-213-95	Sequence 95, Appl
257	30	14.2	52	10	US-09-864-761-35037	Sequence 35037, A	330	29.5	13.9	77	9	US-09-879-572A-30	Sequence 30, Appl
258	30	14.2	52	10	US-09-864-761-41323	Sequence 41323, A	331	29.5	13.9	78	9	US-10-013-379-54	Sequence 54, Appl
259	30	14.2	55	9	US-09-158-722-42	Sequence 42, Appl	332	29.5	13.9	79	9	US-09-874-879-403	Sequence 403, App
260	30	14.2	55	10	US-09-864-761-44800	Sequence 44800, A	333	29.5	13.9	79	9	US-09-813-153-294	Sequence 294, App
261	30	14.2	56	10	US-09-864-761-45250	Sequence 45250, A	334	29.5	13.9	83	10	US-09-764-877-1549	Sequence 1549, Ap
262	30	14.2	58	10	US-09-864-761-48321	Sequence 48321, A	335	29.5	13.9	84	10	US-09-864-761-36104	Sequence 36104, A
263	30	14.2	59	9	US-09-974-979-162	Sequence 162, App	336	29.5	13.9	85	9	US-09-864-761-45026	Sequence 45026, A
264	30	14.2	59	10	US-09-864-761-39239	Sequence 39239, A	337	29.5	13.9	86	10	US-10-114-893-93	Sequence 93, Appl
265	30	14.2	61	10	US-09-864-761-33839	Sequence 33839, A	338	29.5	13.9	86	10	US-09-764-869-1063	Sequence 1063, Ap
266	30	14.2	63	10	US-09-864-761-38772	Sequence 38772, A	339	29.5	13.9	87	9	US-10-016-157A-211	Sequence 211, App
267	30	14.2	63	10	US-09-864-761-48895	Sequence 48895, A	340	29.5	13.9	88	10	US-09-925-297-588	Sequence 588, App
268	30	14.2	64	10	US-09-864-761-46549	Sequence 46549, A	341	29.5	13.9	89	10	US-09-867-550-126	Sequence 126, App
269	30	14.2	65	9	US-09-738-626-6312	Sequence 6312, Ap	342	29	13.7	90	10	US-09-870-962-4	Sequence 4, Appl
270	30	14.2	65	9	US-09-866-050A-237	Sequence 237, App	343	29	13.7	40	9	US-09-738-626-4805	Sequence 4805, Ap
271	30	14.2	65	10	US-09-911-826A-20	Sequence 20, Appl	344	29	13.7	40	10	US-09-864-761-40445	Sequence 40445, A
272	30	14.2	66	10	US-09-864-761-38695	Sequence 38695, A	345	29	13.7	42	9	US-09-925-299-1337	Sequence 1337, App
273	30	14.2	66	10	US-09-864-761-48853	Sequence 48853, A	346	29	13.7	42	9	US-10-020-256A-177	Sequence 177, App
274	30	14.2	67	10	US-09-864-761-48290	Sequence 48290, A	347	29	13.7	42	10	US-09-864-761-35257	Sequence 35257, A
275	30	14.2	67	10	US-09-867-550-1694	Sequence 1694, Ap	348	29	13.7	42	10	US-09-864-761-37467	Sequence 37467, A
276	30	14.2	68	10	US-09-864-761-41976	Sequence 41976, A	349	29	13.7	42	10	US-09-864-761-45391	Sequence 45391, A
277	30	14.2	68	10	US-09-864-761-46121	Sequence 46121, A	350	29	13.7	42	10	US-09-864-761-45911	Sequence 45911, A
278	30	14.2	69	9	US-09-738-626-6750	Sequence 6750, Ap	351	29	13.7	42	10	US-09-864-761-47923	Sequence 47923, A
279	30	14.2	69	12	US-10-001-870-157	Sequence 157, App	352	29	13.7	42	10	US-09-925-299-1337	Sequence 1337, Ap
280	30	14.2	71	9	US-09-738-626-5413	Sequence 5413, App	353	29	13.7	43	10	US-09-864-761-44922	Sequence 44922, A
281	30	14.2	71	9	US-10-083-357-922	Sequence 922, App	354	29	13.7	43	10	US-09-864-761-47736	Sequence 47736, A
282	30	14.2	71	9	US-09-867-361-788	Sequence 788, App	355	29	13.7	44	10	US-09-864-761-34972	Sequence 34972, A
283	30	14.2	72	9	US-09-864-761-36657	Sequence 634, App	356	29	13.7	44	10	US-09-864-761-46624	Sequence 46624, A
284	30	14.2	72	9	US-10-092-154-634	Sequence 724, App	357	29	13.7	45	9	US-09-736-968A-44	Sequence 44, Appl
285	30	14.2	72	10	US-09-764-847-634	Sequence 634, App	358	29	13.7	45	9	US-09-864-761-42757	Sequence 42757, A
286	30	14.2	73	10	US-09-864-761-41902	Sequence 41902, A	359	29	13.7	45	10	US-09-736-969A-44	Sequence 44, Appl
287	30	14.2	73	10	US-09-864-761-48326	Sequence 48326, A	360	29	13.7	45	10	US-09-736-969A-46	Sequence 46, Appl
288	30	14.2	74	10	US-09-917-340-71	Sequence 71, Appl	361	29	13.7	45	10	US-09-939-980-274	Sequence 274, App
289	30	14.2	75	9	US-09-925-299-788	Sequence 788, App	362	29	13.7	45	10	US-09-939-980-274	Sequence 44, Appl
290	30	14.2	75	10	US-09-925-299-788	Sequence 788, App	363	29	13.7	45	10	US-09-736-960-44	Sequence 44, Appl
291	30	14.2	77	10	US-09-864-761-41002	Sequence 41002, A	364	29	13.7	45	10	US-09-736-960-46	Sequence 46, Appl
292	30	14.2	77	10	US-09-864-761-42492	Sequence 42492, A	365	29	13.7	46	10	US-09-864-761-40953	Sequence 40953, A
293	30	14.2	78	9	US-09-736-692-2390	Sequence 2390, Ap	366	29	13.7	46	10	US-09-864-761-48143	Sequence 48143, A
294	30	14.2	78	9	US-09-909-567B-41	Sequence 41, Appl	367	29	13.7	46	10	US-09-779-451-60	Sequence 45, Appl
295	30	14.2	78	9	US-10-092-154-706	Sequence 706, App	368	29	13.7	46	10	US-09-779-451-60	Sequence 1, Appl
296	30	14.2	79	10	US-09-764-847-706	Sequence 706, App	369	29	13.7	46	10	US-09-973-404-1	Sequence 37964, A
297	30	14.2	79	10	US-09-764-887-249	Sequence 249, App	370	29	13.7	47	10	US-09-864-761-47964	Sequence 3657, App
298	30	14.2	81	10	US-09-864-761-34936	Sequence 34936, A	371	29	13.7	47	10	US-09-939-980-365	Sequence 38166, A
299	30	14.2	82	10	US-09-864-761-44395	Sequence 44395, A	372	29	13.7	48	10	US-09-864-761-38166	Sequence 48227, A
300	30	14.2	83	10	US-09-867-550-758	Sequence 758, App	373	29	13.7	48	10	US-09-864-761-48433	Sequence 48433, A
301	30	14.2	85	10	US-09-867-550-2060	Sequence 2060, Ap	374	29	13.7	48	10	US-09-864-761-44834	Sequence 44834, A
302	30	14.2	89	10	US-09-864-761-38560	Sequence 38560, A	375	29	13.7	49	10	US-09-864-761-44834	Sequence 45441, A
303	30	14.2	89	10	US-09-864-761-46552	Sequence 46552, A	376	29	13.7	49	10	US-09-864-761-45441	Sequence 42708, A
304	30	14.2	90	9	US-10-013-477-16	Sequence 16, Appl	377	29	13.7	51	10	US-09-864-761-42708	Sequence 1519, Ap
305	29.5	13.9	42	10	US-09-864-761-42139	Sequence 42139, A	378	29	13.7	52	9	US-09-925-299-1519	Sequence 1519, Ap
306	29.5	13.9	46	10	US-09-864-761-34082	Sequence 34082, A	379	29	13.7	52	10	US-09-864-761-33438	Sequence 33438, A
307	29.5	13.9	46	10	US-09-864-761-42828	Sequence 42828, A	380	29	13.7	52	10	US-09-925-299-1519	Sequence 1519, Ap
308	29.5	13.9	46	10	US-09-864-761-42971	Sequence 42971, A	381	29	13.7	53	9	US-10-043-452-12	Sequence 12, Appl
309	29.5	13.9	48	9	US-09-738-626-4057	Sequence 4057, Ap	382	29	13.7	53	9	US-10-101-464A-831	Sequence 831, App
310	29.5	13.9	49	9	US-10-178-213-93	Sequence 93, Appl	383	29	13.7	53	10	US-09-864-761-41458	Sequence 41458, A
311	29.5	13.9	49	9	US-10-178-213-96	Sequence 96, Appl	384	29	13.7	53	10	US-09-935-428A-21	Sequence 21, Appl

385	29	13.7	53	12	US-10-078-929-190	Sequence 190, App	458	29	13.7	90	10	US-09-818-247-17	Sequence 17, Appl
386	29	13.7	54	10	US-09-864-761-35532	Sequence 35532, A	459	29	13.7	90	10	US-09-751-1008-49	Sequence 49, Appl
387	29	13.7	55	10	US-09-864-761-37370	Sequence 37370, A	460	28.5	13.4	42	10	US-09-864-761-38507	Sequence 38507, A
388	29	13.7	56	10	US-09-864-761-38107	Sequence 38107, A	461	28.5	13.4	49	10	US-09-864-761-47113	Sequence 47113, A
389	29	13.7	57	10	US-09-925-299-1150	Sequence 1150, Ap	462	28.5	13.4	52	9	US-09-764-872-246	Sequence 246, App
390	29	13.7	58	10	US-09-925-299-1150	Sequence 1150, Ap	463	28.5	13.4	52	10	US-09-864-761-34294	Sequence 34294, A
391	29	13.7	59	10	US-09-986-480-348	Sequence 348, App	464	28.5	13.4	53	10	US-09-864-761-36417	Sequence 36417, A
392	29	13.7	60	10	US-09-864-761-41909	Sequence 41909, A	465	28.5	13.4	53	10	US-09-939-980-460	Sequence 460, App
393	29	13.7	61	10	US-09-864-761-42977	Sequence 42977, A	466	28.5	13.4	57	10	US-09-764-870-415	Sequence 415, App
394	29	13.7	62	10	Sequence 42007, A	Sequence 42007, A	467	28.5	13.4	59	10	US-09-205-658-245	Sequence 245, App
395	29	13.7	63	9	US-10-036-692-2400	Sequence 2400, Ap	468	28.5	13.4	61	10	US-09-841-133-6	Sequence 6, Appl
396	29	13.7	64	10	US-09-937-357-1209	Sequence 357, Ap	469	28.5	13.4	62	9	US-09-796-692-1830	Sequence 1830, Ap
397	29	13.7	65	10	US-09-864-761-38739	Sequence 38739, A	470	28.5	13.4	62	9	US-09-796-692-2086	Sequence 2086, Ap
398	29	13.7	66	10	US-09-864-761-44331	Sequence 44331, A	471	28.5	13.4	62	9	US-09-796-692-2316	Sequence 2316, Ap
399	29	13.7	67	10	US-09-864-761-44773	Sequence 44773, A	472	28.5	13.4	62	10	US-09-864-761-40617	Sequence 40617, A
400	29	13.7	68	10	US-09-864-761-36505	Sequence 36505, A	473	28.5	13.4	62	10	US-09-764-877-1735	Sequence 1735, App
401	29	13.7	69	10	US-09-864-761-36919	Sequence 36919, A	474	28.5	13.4	62	10	US-09-864-761-46418	Sequence 46418, A
402	29	13.7	70	10	US-09-864-761-365174	Sequence 45374, A	475	28.5	13.4	64	10	US-09-738-626-3758	Sequence 3758, App
403	29	13.7	71	10	US-09-912-020-357	Sequence 3574, A	476	28.5	13.4	67	9	US-09-841-132-5	Sequence 5, Appl
404	29	13.7	72	10	US-09-864-761-36999	Sequence 36999, App	477	28.5	13.4	67	9	US-09-101-464A-635	Sequence 635, App
405	29	13.7	73	10	US-09-864-761-44665	Sequence 44665, A	478	28.5	13.4	67	10	US-09-864-761-41235	Sequence 41235, A
406	29	13.7	74	10	US-09-731-872-355	Sequence 335, App	479	28.5	13.4	67	10	US-09-864-761-42134	Sequence 42134, A
407	29	13.7	75	10	US-09-925-301-3255	Sequence 38123, A	480	28.5	13.4	72	8	US-08-424-5508-233	Sequence 233, App
408	29	13.7	76	10	US-09-864-761-38123	Sequence 1255, Ap	481	28.5	13.4	72	10	US-09-925-297-569	Sequence 569, App
409	29	13.7	77	10	US-09-871-874-18	Sequence 18, Appl	482	28.5	13.4	73	10	US-09-864-761-48180	Sequence 48180, A
410	29	13.7	78	10	US-09-983-802-277	Sequence 277, App	483	28.5	13.4	75	10	US-09-764-877-1499	Sequence 1499, A
411	29	13.7	79	10	US-09-864-761-35089	Sequence 35089, A	484	28.5	13.4	77	10	US-09-864-761-35099	Sequence 35099, A
412	29	13.7	80	10	US-09-864-761-36677	Sequence 36677, A	485	28.5	13.4	77	10	US-09-864-761-41250	Sequence 41250, A
413	29	13.7	81	10	US-09-864-761-40058	Sequence 40058, A	486	28.5	13.4	78	10	US-09-764-877-1886	Sequence 1886, A
414	29	13.7	82	10	US-09-911-888-353	Sequence 33, Appl	487	28.5	13.4	80	10	US-09-864-761-37250	Sequence 37250, A
415	29	13.7	83	10	US-09-864-761-48665	Sequence 48665, A	488	28.5	13.4	81	10	US-09-864-761-45835	Sequence 45835, A
416	29	13.7	84	10	US-09-045-1701-88	Sequence 88, Appl	489	28.5	13.4	81	10	US-09-764-877-1547	Sequence 1547, A
417	29	13.7	85	10	US-09-738-626-387	Sequence 5287, Ap	490	28.5	13.4	85	10	US-09-764-860-503	Sequence 503, App
418	29	13.7	86	10	US-09-726-626-387	Sequence 88, Appl	491	28.5	13.4	86	10	US-09-841-132-5	Sequence 5, Appl
419	29	13.7	87	10	US-09-864-761-48180	Sequence 48180, A	492	28.5	13.4	88	9	US-09-796-692-2288	Sequence 2288, App
420	29	13.7	88	10	US-09-077-1115-16777	Sequence 46777, A	493	28.5	13.4	89	9	US-10-043-487-521	Sequence 521, App
421	29	13.7	89	10	US-09-867-550-368	Sequence 9, Appl	494	28.5	13.4	90	9	US-09-922-198A-22	Sequence 22, App
422	29	13.7	90	10	US-09-867-550-368	Sequence 968, App	495	28.5	13.4	90	9	US-09-738-626-3796	Sequence 3796, App
423	29	13.7	91	10	US-09-864-761-3518	Sequence 385, App	496	28.5	13.4	90	9	US-10-101-464A-92	Sequence 92, App
424	29	13.7	92	10	US-09-864-761-33523	Sequence 42518, A	497	28	13.2	41	8	US-08-973-601-23	Sequence 23, Appl
425	29	13.7	93	10	US-09-864-761-33523	Sequence 33523, A	498	28	13.2	41	9	US-09-792-630-76	Sequence 76, Appl
426	29	13.7	94	10	US-09-864-761-33523	Sequence 40141, A	499	28	13.2	41	9	US-10-061-395-24	Sequence 24, Appl
427	29	13.7	95	10	US-09-864-921-17041	Sequence 170, App	500	28	13.2	41	9	US-10-096-330-23	Sequence 23, Appl
428	29	13.7	96	10	US-09-815-242-4896	Sequence 4896, App	501	28	13.2	41	10	US-09-157-748-29	Sequence 29, Appl
429	29	13.7	97	10	US-09-862-762-46115	Sequence 46115, A	502	28	13.2	41	10	US-09-864-761-35263	Sequence 35263, A
430	29	13.7	98	10	US-09-862-762-46115	Sequence 46115, A	503	28	13.2	41	10	US-09-864-761-43372	Sequence 43372, A
431	29	13.7	99	10	US-09-862-762-46115	Sequence 1920, Ap	504	28	13.2	41	10	US-09-915-940-27	Sequence 27, Appl
432	29	13.7	100	10	US-09-864-761-37088	Sequence 37088, A	505	28	13.2	41	10	US-09-843-845-8	Sequence 8, Appl
433	29	13.7	101	10	US-09-864-761-37088	Sequence 1336, Ap	506	28	13.2	43	10	US-09-843-845-8	Sequence 8, Appl
434	29	13.7	102	10	US-09-067-065-506	Sequence 506, App	507	28	13.2	43	10	US-09-843-845-22	Sequence 22, Appl
435	29	13.7	103	10	US-09-864-761-35149	Sequence 35149, A	508	28	13.2	43	10	US-09-864-761-41092	Sequence 41092, A
436	29	13.7	104	10	US-09-864-761-45755	Sequence 45755, A	509	28	13.2	44	9	US-09-996-140-165	Sequence 165, App
437	29	13.7	105	10	US-10-125-258-6	Sequence 6, Appl	510	28	13.2	44	9	US-09-996-140-165	Sequence 166, App
438	29	13.7	106	10	US-09-910-150-22	Sequence 22, Appl	511	28	13.2	44	9	US-09-996-140-165	Sequence 167, App
439	29	13.7	107	10	US-09-815-242-10863	Sequence 10863, A	512	28	13.2	44	9	US-09-996-140-165	Sequence 237, App
440	29	13.7	108	10	US-09-796-692-1135	Sequence 1135, Ap.	513	28	13.2	44	9	US-09-996-140-238	Sequence 238, App
441	29	13.7	109	10	US-09-796-692-1585	Sequence 1585, Ap.	514	28	13.2	44	9	US-09-996-140-238	Sequence 239, App
442	29	13.7	110	10	US-09-796-692-2202	Sequence 2202, Ap	515	28	13.2	44	9	US-09-774-639-218	Sequence 218, App
443	29	13.7	111	10	US-09-864-761-35480	Sequence 35480, A	516	28	13.2	44	10	US-09-969-730-371	Sequence 371, Appl
444	29	13.7	112	10	US-09-864-761-44293	Sequence 44293, A	517	28	13.2	44	10	US-09-879-666-7	Sequence 7, Appl
445	29	13.7	113	10	US-09-754-997A-16	Sequence 16, Appl	518	28	13.2	45	10	US-10-042-417-76	Sequence 76, Appl
446	29	13.7	114	10	US-09-808-387-34	Sequence 34, Appl	519	28	13.2	45	10	US-10-178-213-15	Sequence 15, Appl
447	29	13.7	115	10	US-10-081-817-22	Sequence 22, Appl	520	28	13.2	45	10	US-09-864-761-47953	Sequence 47953, A
448	29	13.7	116	10	US-09-864-761-46848	Sequence 46848, A	521	28	13.2	45	10	US-09-925-300-946	Sequence 946, App
449	29	13.7	117	10	US-10-081-817-2	Sequence 2, Appl	522	28	13.2	46	10	US-09-813-153-201	Sequence 201, App
450	29	13.7	118	10	US-09-864-761-41682	Sequence 193, App	523	28	13.2	46	10	US-09-864-761-37981	Sequence 37981, A
451	29	13.7	119	10	US-09-867-550-462	Sequence 462, App	524	28	13.2	46	10	US-09-864-761-48128	Sequence 48128, A
452	29	13.7	120	10	US-09-764-878-193	Sequence 193, App	525	28	13.2	46	10	US-09-864-761-48604	Sequence 48604, A
453	29	13.7	121	9	US-09-852-797-98	Sequence 98, Appl	526	28	13.2	46	10	US-09-854-816-109	Sequence 109, App
454	29	13.7	122	10	US-09-738-626-4015	Sequence 4015, Ap	527	28	13.2	47	10	US-09-864-761-34601	Sequence 34601, A
455	29	13.7	123	10	US-09-853-161-98	Sequence 98, Appl	528	28	13.2	47	10	US-09-764-869-743	Sequence 743, App
456	29	13.7	124	10	US-09-832-659A-98	Sequence 98, Appl	529	28	13.2	47	10	US-09-848-664-26	Sequence 26, Appl
457	29	13.7	125	10	US-09-925-297-711	Sequence 711, Appl	530	28	13.2	48	10	US-09-729-674-46	Sequence 46, Appl
458	29	13.7	126	9	US-09-738-626-6089	Sequence 6089, App	531	28	13.2	48	10	US-09-864-761-39565	Sequence 39565, A
459	29	13.7	127	9	US-10-100-252-6	Sequence 6, Appl	532	28	13.2	48	10	US-09-864-761-39565	Sequence 39565, A



531	28	13.2	48	10	US-09-864-761-41071	Sequence 41071, A	604	28	13.2	75	9	US-09-978-303-17	Sequence 17, Appl
532	28	13.2	49	9	US-10-023-066A-30	Sequence 30, Appl	605	28	13.2	75	10	US-09-864-761-34641	Sequence 34641, A
533	28	13.2	49	9	US-10-023-066A-32	Sequence 32, Appl	606	28	13.2	75	10	US-09-939-980-415	Sequence 415, App
534	28	13.2	49	9	US-10-023-066A-54	Sequence 54, Appl	607	28	13.2	76	9	US-09-749-637A-207	Sequence 207, App
535	28	13.2	49	9	US-09-996-140-164	Sequence 164, App	608	28	13.2	76	10	US-09-864-761-34140	Sequence 34140, A
536	28	13.2	50	9	US-09-996-140-236	Sequence 236, App	609	28	13.2	76	10	US-09-864-761-39315	Sequence 39315, A
537	28	13.2	50	9	US-10-016-634A-139	Sequence 139, App	610	28	13.2	76	10	US-09-864-761-41894	Sequence 41894, A
538	28	13.2	50	9	US-09-796-692-1810	Sequence 1810, App	611	28	13.2	76	10	US-09-864-761-44615	Sequence 44615, A
539	28	13.2	50	10	US-09-823-266-35	Sequence 25, Appl	612	28	13.2	77	9	US-09-872-462-6	Sequence 6, Appl
540	28	13.2	51	10	US-09-864-761-35785	Sequence 35785, A	613	28	13.2	77	10	US-09-864-761-40148	Sequence 40148, A
541	28	13.2	52	9	US-09-730-763-10	Sequence 10, Appl	614	28	13.2	77	10	US-09-864-761-47322	Sequence 47322, A
542	28	13.2	52	9	US-09-866-050A-167	Sequence 167, App	615	28	13.2	77	10	US-09-867-550-242	Sequence 242, App
543	28	13.2	52	9	US-10-079-854-147	Sequence 147, App	616	28	13.2	77	10	US-09-764-860-372	Sequence 372, App
544	28	13.2	52	10	US-09-864-761-34375	Sequence 34375, A	617	28	13.2	78	9	US-09-738-626-5562	Sequence 5562, App
545	28	13.2	52	10	US-09-864-761-46407	Sequence 46407, A	618	28	13.2	78	9	US-09-796-692-2426	Sequence 2426, App
546	28	13.2	52	10	US-09-764-878-147	Sequence 147, App	619	28	13.2	78	10	US-09-864-761-33969	Sequence 33969, A
547	28	13.2	53	9	US-09-796-692-2455	Sequence 2455, App	620	28	13.2	78	10	US-09-864-761-37973	Sequence 37973, A
548	28	13.2	53	10	US-09-864-761-35091	Sequence 35091, A	621	28	13.2	78	10	US-09-815-242-10391	Sequence 10391, A
549	28	13.2	53	10	US-09-864-761-44587	Sequence 44587, A	622	28	13.2	78	12	US-09-738-626-5545	Sequence 5545, App
550	28	13.2	54	10	US-09-864-761-34369	Sequence 34369, A	623	28	13.2	79	10	US-09-864-761-47322	Sequence 47322, A
551	28	13.2	54	10	US-09-864-761-40709	Sequence 40709, A	624	28	13.2	79	10	US-09-864-761-47317	Sequence 47317, A
552	28	13.2	54	10	US-09-864-761-42322	Sequence 42322, A	625	28	13.2	80	9	US-09-925-299-1244	Sequence 1244, App
553	28	13.2	54	10	US-09-864-761-40066	Sequence 40066, A	626	28	13.2	80	9	US-10-052-154-626	Sequence 626, App
554	28	13.2	55	10	US-09-939-980-414	Sequence 414, App	627	28	13.2	80	9	US-10-052-154-626	Sequence 1244, App
555	28	13.2	55	10	US-10-023-066A-79	Sequence 79, Appl	628	28	13.2	80	10	US-09-925-299-1244	Sequence 626, App
556	28	13.2	56	10	US-09-864-761-41588	Sequence 41588, A	629	28	13.2	80	10	US-09-764-847-6326	Sequence 1244, App
557	28	13.2	56	10	US-09-779-451-4	Sequence 4, Appl	630	28	13.2	81	9	US-09-924-340-38	Sequence 38, Appl
558	28	13.2	56	10	US-09-954-737-3	Sequence 3, Appl	631	28	13.2	81	9	US-09-992-600A-38	Sequence 38, Appl
559	28	13.2	57	10	US-09-864-761-39078	Sequence 39078, A	632	28	13.2	81	10	US-09-864-761-46595	Sequence 46595, A
560	28	13.2	58	10	US-09-809-545A-55	Sequence 55, Appl	633	28	13.2	82	9	US-09-983-802-592	Sequence 592, App
561	28	13.2	58	10	US-10-001-835-205	Sequence 205, App	634	28	13.2	82	10	US-09-815-242-13871	Sequence 13871, A
562	28	13.2	59	10	US-09-864-761-40530	Sequence 40530, A	635	28	13.2	83	9	US-10-083-357-792	Sequence 792, App
563	28	13.2	60	10	US-09-864-761-34440	Sequence 34440, A	636	28	13.2	84	9	US-09-925-299-1207	Sequence 1207, App
564	28	13.2	62	9	US-09-738-626-5087	Sequence 5087, App	637	28	13.2	84	10	US-09-864-761-45989	Sequence 45989, A
565	28	13.2	62	9	US-10-031-572-368	Sequence 368, Appl	638	28	13.2	84	10	US-09-925-299-1207	Sequence 1207, App
566	28	13.2	62	10	US-09-967-552A-22	Sequence 22, Appl	639	28	13.2	84	10	US-09-925-299-1207	Sequence 1166, App
567	28	13.2	63	10	US-09-967-552A-58	Sequence 58, Appl	640	28	13.2	85	10	US-09-864-761-38322	Sequence 38322, A
568	28	13.2	63	10	US-09-780-264-4	Sequence 4, Appl	641	28	13.2	85	10	US-09-864-761-38883	Sequence 42883, A
569	28	13.2	63	10	US-09-864-761-38406	Sequence 38406, A	642	28	13.2	86	9	US-10-083-357-1070	Sequence 1070, App
570	28	13.2	63	10	US-09-864-761-46403	Sequence 46403, A	643	28	13.2	87	10	US-09-764-860-347	Sequence 147, App
571	28	13.2	64	10	US-09-757-908A-2	Sequence 2, Appl	644	28	13.2	88	10	US-09-811-284-194	Sequence 194, App
572	28	13.2	64	10	US-09-864-761-35626	Sequence 35626, A	645	28	13.2	88	10	US-09-815-242-5525	Sequence 5525, App
573	28	13.2	65	8	US-08-808-031A-18	Sequence 18, Appl	646	28	13.2	88	10	US-09-815-242-12350	Sequence 12350, A
574	28	13.2	65	9	US-09-796-692-1993	Sequence 1993, App	647	28	13.2	88	10	US-09-815-242-12932	Sequence 12932, A
575	28	13.2	65	9	US-09-796-692-2232	Sequence 2232, App	648	28	13.2	88	10	US-09-815-242-13147	Sequence 13147, A
576	28	13.2	65	9	US-10-001-254-53	Sequence 53, Appl	649	28	13.2	88	10	US-09-815-242-13449	Sequence 13449, A
577	28	13.2	66	10	US-09-967-552A-60	Sequence 60, Appl	650	28	13.2	88	10	US-09-815-242-13694	Sequence 13694, A
578	28	13.2	67	9	US-09-738-626-3845	Sequence 3845, App	651	28	13.2	89	9	US-10-002-344A-156	Sequence 156, App
579	28	13.2	67	10	US-09-864-761-44418	Sequence 44418, A	652	28	13.2	89	9	US-10-001-876-193	Sequence 193, App
580	28	13.2	67	10	US-09-864-761-48460	Sequence 48460, A	653	28	13.2	89	9	US-10-037-065-630	Sequence 630, App
581	28	13.2	67	10	US-09-925-297-561	Sequence 561, App	654	28	13.2	89	10	US-09-815-242-5353	Sequence 5353, App
582	28	13.2	68	10	US-09-864-761-38568	Sequence 38568, A	655	28	13.2	89	10	US-09-815-242-13211	Sequence 13211, A
583	28	13.2	68	10	US-09-864-761-45307	Sequence 45307, A	656	28	13.2	89	10	US-09-840-459-50	Sequence 50, Appl
584	28	13.2	68	10	US-09-925-301-1492	Sequence 1492, App	657	28	13.2	90	9	US-09-976-059-12	Sequence 12, Appl
585	28	13.2	69	9	US-10-037-065-227	Sequence 227, App	658	28	13.2	90	9	US-09-738-626-3545	Sequence 3545, App
586	28	13.2	69	10	US-09-864-761-40598	Sequence 40598, A	659	28	13.2	90	9	US-09-925-299-1529	Sequence 1529, App
587	28	13.2	70	10	US-09-864-761-35373	Sequence 35373, A	660	28	13.2	90	10	US-09-925-299-1529	Sequence 1529, App
588	28	13.2	70	10	US-09-864-761-40956	Sequence 40956, A	661	28	13.2	90	10	US-09-876-889-17	Sequence 17, Appl
589	28	13.2	70	10	US-09-880-192-49	Sequence 49, Appl	662	28	13.2	90	10	US-10-016-157A-228	Sequence 228, App
590	28	13.2	70	10	US-09-867-550-1774	Sequence 1774, App	663	27.5	13.0	47	9	US-09-864-761-33481	Sequence 33481, A
591	28	13.2	71	10	US-09-864-761-33717	Sequence 33717, App	664	27.5	13.0	47	10	US-09-764-877-1115	Sequence 1115, App
592	28	13.2	71	10	US-09-864-761-35926	Sequence 35926, A	665	27.5	13.0	48	10	US-09-864-761-45980	Sequence 45980, A
593	28	13.2	71	10	US-09-764-877-1688	Sequence 1688, App	666	27.5	13.0	50	10	US-09-864-761-43198	Sequence 43198, A
594	28	13.2	73	1	US-08-954-771-35	Sequence 35, Appl	667	27.5	13.0	51	10	US-09-864-761-43821	Sequence 43821, A
595	28	13.2	73	1	US-08-954-771-36	Sequence 36, Appl	668	27.5	13.0	51	10	US-09-864-761-42875	Sequence 42875, A
596	28	13.2	74	9	US-09-796-692-1973	Sequence 1973, App	669	27.5	13.0	53	10	US-09-864-761-1983	Sequence 1983, App
597	28	13.2	74	9	US-09-796-692-2489	Sequence 2489, App	670	27.5	13.0	54	10	US-09-864-761-40033	Sequence 40033, A
598	28	13.2	74	10	US-09-741-669-387	Sequence 387, App	671	27.5	13.0	54	10	US-09-854-816-75	Sequence 75, Appl
599	28	13.2	74	10	US-09-864-761-33496	Sequence 33496, A	672	27.5	13.0	55	9	US-09-738-626-4735	Sequence 4735, App
600	28	13.2	74	10	US-09-864-761-42540	Sequence 42540, A	673	27.5	13.0	55	10	US-09-864-761-39961	Sequence 39961, A
601	28	13.2	74	10	US-09-819-058-2	Sequence 2, Appl	674	27.5	13.0	55	10	US-09-864-761-41516	Sequence 41516, A
602	28	13.2	75	8	US-08-424-550B-407	Sequence 407, App	675	27.5	13.0	56	10	US-09-864-761-36664	Sequence 36664, A
603	28	13.2	75	9	US-09-978-303-16	Sequence 16, Appl	676	27.5	13.0	61	9	US-09-738-626-6564	Sequence 6564, App

677	27.5	13.0	61	9	US-09-796-692-829	Sequence 829, App	750	27	12.7	49	10	US-09-873-637-18	Sequence 18, Appl
678	27.5	13.0	61	9	US-09-796-692-829	Sequence 2419, App	751	27	12.7	50	9	US-10-077-111-24	Sequence 24, Appl
679	27.5	13.0	61	9	US-09-867-550-2066	Sequence 2066, App	752	27	12.7	50	10	US-09-864-761-34397	Sequence 34397, A
680	27.5	13.0	62	10	US-09-855-341-15	Sequence 15, Appl	753	27	12.7	50	10	US-09-864-761-34616	Sequence 34616, A
681	27.5	13.0	62	10	US-09-864-761-34486	Sequence 34486, A	754	27	12.7	50	10	US-09-864-761-38825	Sequence 38825, A
682	27.5	13.0	62	10	US-09-864-761-35320	Sequence 35320, A	755	27	12.7	50	10	US-09-864-761-38825	Sequence 38825, A
683	27.5	13.0	63	9	US-10-012-542-449	Sequence 449, App	756	27	12.7	50	10	US-09-864-761-45255	Sequence 45255, A
684	27.5	13.0	64	9	US-09-870-759-116	Sequence 116, App	757	27	12.7	50	10	US-09-864-761-47575	Sequence 47575, A
685	27.5	13.0	64	10	US-09-864-761-39069	Sequence 39069, A	758	27	12.7	50	10	US-09-764-869-879	Sequence 879, App
686	27.5	13.0	67	10	US-09-815-242-13059	Sequence 13059, A	759	27	12.7	51	9	US-09-774-638-214	Sequence 214, App
687	27.5	13.0	69	9	US-09-738-626-6225	Sequence 6222, App	760	27	12.7	51	10	US-09-663-730-222	Sequence 222, App
688	27.5	13.0	70	10	US-09-864-761-37767	Sequence 37767, A	761	27	12.7	51	12	US-09-864-761-45585	Sequence 45585, A
689	27.5	13.0	71	10	US-09-864-761-36525	Sequence 36525, A	762	27	12.7	51	12	US-10-124-557-100	Sequence 100, App
690	27.5	13.0	72	9	US-09-809-351-566	Sequence 566, App	763	27	12.7	52	12	US-10-124-557-112	Sequence 112, App
691	27.5	13.0	73	10	US-09-864-761-44645	Sequence 44645, A	764	27	12.7	52	10	US-09-728-912-12	Sequence 12, Appl
692	27.5	13.0	73	10	US-09-864-761-42885	Sequence 42885, A	765	27	12.7	53	9	US-09-843-676-20	Sequence 20, Appl
693	27.5	13.0	74	9	US-09-738-626-4360	Sequence 4360, App	766	27	12.7	53	9	US-09-766-253-20	Sequence 20, Appl
694	27.5	13.0	75	10	US-09-864-761-46901	Sequence 46901, A	767	27	12.7	53	9	US-10-016-157A-229	Sequence 229, App
695	27.5	13.0	76	10	US-09-864-761-47518	Sequence 47518, A	768	27	12.7	53	9	US-09-438-486-20	Sequence 20, Appl
696	27.5	13.0	77	10	US-09-867-550-1432	Sequence 1432, App	769	27	12.7	53	9	US-10-053-758-20	Sequence 20, Appl
697	27.5	13.0	80	10	US-09-767-007A-3	Sequence 3, Appl	770	27	12.7	53	9	US-10-054-295-20	Sequence 20, Appl
698	27.5	13.0	81	9	US-09-753-126-3	Sequence 519, App	771	27	12.7	53	10	US-09-864-761-41983	Sequence 41983, A
699	27.5	13.0	82	10	US-09-867-550-550	Sequence 550, App	772	27	12.7	53	10	US-09-864-761-42900	Sequence 42900, A
700	27.5	13.0	82	10	US-09-764-864-1460	Sequence 1460, App	773	27	12.7	54	9	US-09-738-626-3868	Sequence 3868, App
701	27.5	13.0	83	9	US-10-091-572-237	Sequence 237, App	774	27	12.7	54	10	US-09-864-761-34308	Sequence 34308, A
702	27.5	13.0	84	10	US-09-864-761-42438	Sequence 42438, A	775	27	12.7	54	10	US-09-864-761-37646	Sequence 37646, A
703	27.5	13.0	84	10	US-09-864-761-43315	Sequence 43315, A	776	27	12.7	54	10	US-09-864-761-42543	Sequence 42543, A
704	27.5	13.0	85	10	US-09-867-550-1432	Sequence 1432, App	777	27	12.7	54	10	US-09-867-550-496	Sequence 496, App
705	27.5	13.0	86	10	US-09-925-300-1139	Sequence 1139, App	778	27	12.7	55	9	US-10-091-572-241	Sequence 241, App
706	27.5	13.0	87	9	US-10-083-357-918	Sequence 918, App	779	27	12.7	55	9	US-09-908-711-80	Sequence 80, Appl
707	27.5	13.0	88	10	US-09-971-536-45	Sequence 45688, A	780	27	12.7	55	10	US-09-864-761-42481	Sequence 42481, A
708	27.5	13.0	88	10	US-09-864-761-46888	Sequence 46888, A	781	27	12.7	55	10	US-09-986-194-17	Sequence 17, Appl
709	27.5	13.0	88	10	US-09-815-242-4943	Sequence 4943, App	782	27	12.7	56	9	US-10-138-082-9	Sequence 9, Appl
710	27.5	13.0	89	10	US-09-815-242-10784	Sequence 10784, A	783	27	12.7	56	9	US-09-788-006-56	Sequence 56, Appl
711	27.5	13.0	89	10	US-10-134-841-1	Sequence 1, Appl	784	27	12.7	56	9	US-10-023-066A-77	Sequence 77, Appl
712	27.5	13.0	89	9	US-09-481-322-10	Sequence 10, Appl	785	27	12.7	56	10	US-09-864-761-34673	Sequence 34673, A
713	27.5	13.0	90	9	US-09-481-322-11	Sequence 11, Appl	786	27	12.7	56	10	US-09-864-761-35747	Sequence 35747, A
714	27.5	13.0	90	10	US-09-864-761-35729	Sequence 35729, A	787	27	12.7	56	10	US-09-864-761-37976	Sequence 37976, A
715	27.5	13.0	90	10	US-09-864-761-35729	Sequence 86, Appl	788	27	12.7	56	10	US-09-864-761-38796	Sequence 38796, A
716	27.5	13.0	91	9	US-09-974-639-208	Sequence 206, App	789	27	12.7	56	10	US-09-864-761-47594	Sequence 47594, A
717	27.5	13.0	91	9	US-09-974-639-208	Sequence 311, App	790	27	12.7	56	10	US-09-777-422-9	Sequence 9, Appl
718	27.5	13.0	91	9	US-09-974-639-208	Sequence 42451, A	791	27	12.7	57	9	US-10-012-896-565	Sequence 565, App
719	27.5	13.0	91	9	US-09-974-639-208	Sequence 11, App	792	27	12.7	57	9	US-09-895-793-565	Sequence 565, App
720	27.5	13.0	91	9	US-09-974-639-208	Sequence 35309, A	793	27	12.7	57	9	US-09-895-814-565	Sequence 565, App
721	27.5	13.0	91	9	US-09-864-761-42451	Sequence 40208, A	794	27	12.7	57	9	US-10-101-464A-830	Sequence 830, App
722	27.5	13.0	92	10	US-09-864-761-35309	Sequence 40355, A	795	27	12.7	57	10	US-09-759-143-565	Sequence 565, App
723	27.5	13.0	92	10	US-09-864-761-40208	Sequence 41495, A	796	27	12.7	57	10	US-09-864-761-42482	Sequence 42482, A
724	27.5	13.0	93	10	US-09-864-761-40355	Sequence 46455, A	797	27	12.7	57	10	US-09-864-761-42649	Sequence 42649, A
725	27.5	13.0	93	10	US-09-864-761-41495	Sequence 466, App	798	27	12.7	57	10	US-09-864-761-42820	Sequence 42820, A
726	27.5	13.0	93	10	US-09-864-761-46455	Sequence 4106, App	799	27	12.7	57	10	US-09-822-827-565	Sequence 565, App
727	27.5	13.0	94	9	US-10-097-065-466	Sequence 4196, A	800	27	12.7	57	10	US-09-734-017A-52	Sequence 52, App
728	27.5	13.0	94	10	US-09-864-761-41106	Sequence 4196, A	801	27	12.7	57	10	US-09-986-480-315	Sequence 315, Appl
729	27.5	13.0	94	10	US-09-864-761-42196	Sequence 139, App	802	27	12.7	58	9	US-10-092-154-53A	Sequence 354, App
730	27.5	13.0	94	9	US-10-072-349-139	Sequence 139, App	803	27	12.7	58	9	US-09-864-761-35606	Sequence 35606, A
731	27.5	13.0	95	9	US-09-728-912-13	Sequence 13, App	804	27	12.7	58	10	US-09-864-761-36239	Sequence 36239, A
732	27.5	13.0	95	10	US-09-864-761-38050	Sequence 38050, A	805	27	12.7	58	10	US-09-864-761-36239	Sequence 23, Appl
733	27.5	13.0	95	10	US-09-864-761-38050	Sequence 41747, A	806	27	12.7	58	10	US-09-948-080-23	Sequence 534, App
734	27.5	13.0	95	10	US-09-864-761-41747	Sequence 44066, A	807	27	12.7	58	10	US-09-764-847-534	Sequence 1038, App
735	27.5	13.0	95	10	US-09-864-761-44066	Sequence 50, Appl	808	27	12.7	59	9	US-09-796-692-1038	Sequence 473, App
736	27.5	13.0	95	10	US-09-923-679-50	Sequence 139, App	809	27	12.7	59	9	US-10-043-487-473	Sequence 37411, A
737	27.5	13.0	95	10	US-09-764-855-139	Sequence 563, App	810	27	12.7	59	10	US-09-864-761-37411	Sequence 312, App
738	27.5	13.0	96	9	US-09-983-802-563	Sequence 39211, A	811	27	12.7	59	10	US-09-864-761-40219	Sequence 40219, A
739	27.5	13.0	96	10	US-09-864-761-39211	Sequence 44259, A	812	27	12.7	60	9	US-09-864-761-40602	Sequence 40602, A
740	27.5	13.0	96	10	US-09-864-761-44259	Sequence 156, App	813	27	12.7	60	9	US-09-864-761-43178	Sequence 43178, A
741	27.5	13.0	97	12	US-10-001-879-156	Sequence 172, App	814	27	12.7	60	9	US-09-864-761-45851	Sequence 45851, A
742	27.5	13.0	97	12	US-09-574-879-172	Sequence 977, App	815	27	12.7	60	10	US-09-864-761-46375	Sequence 46375, A
743	27.5	13.0	97	12	US-09-574-879-172	Sequence 1780, A	816	27	12.7	60	10	US-10-001-843-191	Sequence 191, App
744	27.5	13.0	98	9	US-10-083-357-977	Sequence 977, App	817	27	12.7	60	10	US-09-764-868-1230	Sequence 1230, App
745	27.5	13.0	98	10	US-09-864-761-35514	Sequence 35514, A	818	27	12.7	60	10		
746	27.5	13.0	98	10	US-09-864-761-36507	Sequence 36507, A	819	27	12.7	60	10		
747	27.5	13.0	99	10	US-09-864-761-36507	Sequence 36507, A	820	27	12.7	60	10		
748	27.5	13.0	99	10	US-09-864-761-36507	Sequence 42993, A	821	27	12.7	60	12		
749	27.5	13.0	99	10	US-09-864-761-42993	Sequence 43231, A	822	27	12.7	61	9		

823	27	12.7	61	9	US-10-091-483-166	Sequence 166, App	896	27	12.7	78	10	US-09-740-288A-16	Sequence 16, Appl
824	27	12.7	61	10	US-09-764-846-166	Sequence 166, App	897	27	12.7	78	10	US-09-864-761-35419	Sequence 35419, A
825	27	12.7	62	9	US-09-309-196-111	Sequence 111, App	898	27	12.7	78	10	US-09-925-299-1065	Sequence 1065, Ap
826	27	12.7	62	9	US-09-309-196-111	Sequence 116, App	899	27	12.7	79	9	US-10-012-542-197	Sequence 197, App
827	27	12.7	62	9	US-09-309-196-111	Sequence 572, App	900	27	12.7	79	9	US-10-092-154-903	Sequence 903, App
828	27	12.7	62	9	US-09-864-761-36344	Sequence 36344, A	901	27	12.7	79	9	US-10-092-154-987	Sequence 987, App
829	27	12.7	62	10	US-09-864-761-39201	Sequence 39201, A	902	27	12.7	79	10	US-09-864-761-40697	Sequence 40697, A
830	27	12.7	62	10	US-09-864-761-42374	Sequence 42374, A	903	27	12.7	79	10	US-09-864-761-48228	Sequence 48228, A
831	27	12.7	62	10	US-09-864-761-42374	Sequence 572, App	904	27	12.7	79	10	US-09-815-243-1316	Sequence 1316, A
832	27	12.7	63	9	US-09-994-595-139	Sequence 139, App	905	27	12.7	79	10	US-09-815-243-13568	Sequence 13568, A
833	27	12.7	63	10	US-09-864-761-34328	Sequence 34328, A	906	27	12.7	79	10	US-09-764-847-903	Sequence 903, App
834	27	12.7	63	10	US-09-864-761-43716	Sequence 43716, A	907	27	12.7	79	10	US-09-764-847-987	Sequence 987, App
835	27	12.7	63	12	US-10-035-408-2	Sequence 2, Appl	908	27	12.7	79	10	US-09-925-300-1401	Sequence 1401, Ap
836	27	12.7	64	9	US-09-809-391-414	Sequence 414, App	909	27	12.7	80	10	US-09-864-761-35189	Sequence 35189, A
837	27	12.7	64	10	US-09-864-761-41845	Sequence 41845, A	910	27	12.7	80	10	US-09-864-761-46053	Sequence 46053, A
838	27	12.7	65	9	US-09-981-876-171	Sequence 171, App	911	27	12.7	80	10	US-09-932-679-54	Sequence 54, Appl
839	27	12.7	65	9	US-09-148-545-171	Sequence 171, App	912	27	12.7	81	9	US-09-730-617-72	Sequence 72, Appl
840	27	12.7	65	10	US-09-764-877-1590	Sequence 1590, Ap	913	27	12.7	81	9	US-09-796-692-2466	Sequence 2466, Ap
841	27	12.7	66	10	US-09-864-761-38266	Sequence 38266, A	914	27	12.7	81	10	US-09-925-300-1172	Sequence 1172, Ap
842	27	12.7	66	10	US-09-864-761-42532	Sequence 42532, A	915	27	12.7	82	10	US-09-764-877-1582	Sequence 1582, Ap
843	27	12.7	67	10	US-09-864-761-36942	Sequence 36942, A	916	27	12.7	83	9	US-10-002-3448-202	Sequence 202, App
844	27	12.7	67	10	US-09-864-761-43986	Sequence 43986, A	917	27	12.7	83	9	US-10-002-3448-202	Sequence 203, App
845	27	12.7	67	10	US-09-867-550-456	Sequence 456, App	918	27	12.7	83	9	US-10-083-357-785	Sequence 785, App
846	27	12.7	67	10	US-09-867-550-814	Sequence 814, App	919	27	12.7	83	10	US-09-131-827A-13	Sequence 13, Appl
847	27	12.7	68	9	US-09-158-722-31	Sequence 31, Appl	920	27	12.7	83	10	US-09-925-297-845	Sequence 845, App
848	27	12.7	68	9	US-10-092-134-929	Sequence 929, App	921	27	12.7	83	10	US-09-867-550-404	Sequence 404, App
849	27	12.7	68	10	US-09-864-761-43727	Sequence 43727, A	922	27	12.7	84	10	US-09-864-761-36633	Sequence 36633, A
850	27	12.7	68	10	US-09-764-847-929	Sequence 929, App	923	27	12.7	84	10	US-09-731-872-330	Sequence 330, App
851	27	12.7	69	9	US-10-083-357-829	Sequence 829, App	924	27	12.7	85	10	US-09-730-617-71	Sequence 71, Appl
852	27	12.7	69	9	US-10-097-065-527	Sequence 527, App	925	27	12.7	85	10	US-09-730-617-73	Sequence 73, Appl
853	27	12.7	69	10	US-09-864-761-39488	Sequence 39488, A	926	27	12.7	86	9	US-10-102-806-556	Sequence 556, App
854	27	12.7	69	10	US-09-864-761-45716	Sequence 45716, A	927	27	12.7	86	10	US-09-864-761-42654	Sequence 42654, A
855	27	12.7	69	10	US-09-864-761-45962	Sequence 45962, A	928	27	12.7	87	9	US-09-738-626-4627	Sequence 4627, Ap
856	27	12.7	69	10	US-09-853-386-72	Sequence 72, Appl	929	27	12.7	88	9	US-10-116-255-34	Sequence 34, Appl
857	27	12.7	70	9	US-10-068-564-12	Sequence 12, Appl	930	27	12.7	88	10	US-09-764-869-903	Sequence 903, App
858	27	12.7	70	10	US-09-864-761-34446	Sequence 34446, A	931	27	12.7	88	10	US-09-354-453-51	Sequence 51, Appl
859	27	12.7	70	10	US-09-864-761-37272	Sequence 37272, A	932	27	12.7	88	10	US-09-843-676-204	Sequence 204, App
860	27	12.7	70	10	US-09-864-761-42794	Sequence 42794, A	933	27	12.7	89	9	US-09-738-626-5203	Sequence 5203, Ap
861	27	12.7	70	10	US-09-903-814-8	Sequence 8, Appl	934	27	12.7	89	9	US-09-438-486-204	Sequence 204, App
862	27	12.7	70	10	US-09-989-903-12	Sequence 12, Appl	935	27	12.7	89	9	US-09-768-235B-12	Sequence 12, Appl
863	27	12.7	70	10	US-09-764-877-1589	Sequence 1589, Ap	936	27	12.7	89	9	US-10-053-758-204	Sequence 204, App
864	27	12.7	70	10	US-09-854-816-102	Sequence 102, App	937	27	12.7	89	9	US-10-054-295-204	Sequence 204, App
865	27	12.7	71	10	US-09-864-761-39931	Sequence 39931, A	938	27	12.7	89	9	US-10-057-065-162	Sequence 162, App
866	27	12.7	72	9	US-09-796-692-865	Sequence 865, App	939	27	12.7	89	10	US-09-729-674-10	Sequence 10, Appl
867	27	12.7	72	10	US-09-864-761-36642	Sequence 36642, A	940	27	12.7	89	10	US-09-864-761-37241	Sequence 37241, A
868	27	12.7	72	10	US-09-864-761-43253	Sequence 43253, A	941	27	12.7	89	10	US-09-864-761-37664	Sequence 37664, A
869	27	12.7	72	10	US-09-854-816-83	Sequence 83, Appl	942	27	12.7	89	10	US-09-734-569-110	Sequence 110, App
870	27	12.7	73	10	US-09-864-761-35727	Sequence 35727, A	943	27	12.7	89	10	US-09-764-877-1746	Sequence 1746, App
871	27	12.7	73	10	US-09-864-761-36723	Sequence 36723, A	944	27	12.7	90	10	US-09-939-980-427	Sequence 427, Appl
872	27	12.7	74	9	US-09-764-868-931	Sequence 931, App	945	26.5	12.5	40	10	US-09-939-980-427	Sequence 427, Appl
873	27	12.7	74	9	US-09-738-628-4017	Sequence 4017, App	946	26.5	12.5	41	10	US-09-864-761-45042	Sequence 45042, A
874	27	12.7	74	9	US-09-738-628-4302	Sequence 4302, App	947	26.5	12.5	42	10	US-09-864-761-44678	Sequence 44678, A
875	27	12.7	74	9	US-09-738-628-4992	Sequence 4992, App	948	26.5	12.5	43	10	US-09-879-957-108	Sequence 108, App
876	27	12.7	74	9	US-09-910-009A-47	Sequence 47, Appl	949	26.5	12.5	43	10	US-09-864-761-43136	Sequence 43136, A
877	27	12.7	74	9	US-09-910-009A-372	Sequence 372, App	950	26.5	12.5	44	10	US-09-843-845-16	Sequence 16, Appl
878	27	12.7	74	10	US-09-110-716-38	Sequence 38, Appl	951	26.5	12.5	44	10	US-09-925-300-1089	Sequence 1089, Ap
879	27	12.7	74	10	US-09-864-761-36478	Sequence 36478, A	952	26.5	12.5	45	9	US-09-736-968A-41	Sequence 41, Appl
880	27	12.7	74	10	US-09-932-679-51	Sequence 51, Appl	953	26.5	12.5	45	9	US-09-736-968A-42	Sequence 42, Appl
881	27	12.7	75	9	US-09-976-059-28	Sequence 28, Appl	954	26.5	12.5	45	10	US-09-736-969A-41	Sequence 41, Appl
882	27	12.7	75	9	US-09-796-692-870	Sequence 870, App	955	26.5	12.5	45	10	US-09-736-969A-42	Sequence 42, Appl
883	27	12.7	75	9	US-09-925-299-1426	Sequence 1426, App	956	26.5	12.5	45	10	US-09-736-960-41	Sequence 41, Appl
884	27	12.7	75	9	US-09-910-009A-11	Sequence 11, Appl	957	26.5	12.5	45	10	US-09-736-960-42	Sequence 42, Appl
885	27	12.7	75	10	US-09-729-674-34	Sequence 34, Appl	958	26.5	12.5	47	10	US-09-864-761-38847	Sequence 38847, A
886	27	12.7	75	10	US-09-864-761-40989	Sequence 40989, A	959	26.5	12.5	47	10	US-09-864-761-43292	Sequence 43292, A
887	27	12.7	75	10	US-09-864-761-47129	Sequence 47129, A	960	26.5	12.5	48	10	US-09-739-254-74	Sequence 74, Appl
888	27	12.7	76	9	US-09-820-843A-52	Sequence 52, Appl	961	26.5	12.5	48	10	US-09-904-615-74	Sequence 74, Appl
889	27	12.7	76	9	US-09-925-299-1496	Sequence 1496, App	962	26.5	12.5	48	10	US-09-864-761-40480	Sequence 40480, A
890	27	12.7	76	9	US-09-994-595-153	Sequence 153, App	963	26.5	12.5	50	9	US-10-012-896-585	Sequence 585, App
891	27	12.7	76	10	US-09-925-299-1496	Sequence 1496, App	964	26.5	12.5	50	9	US-09-895-793-585	Sequence 585, App
892	27	12.7	76	10	US-09-993-844-49	Sequence 49, Appl	965	26.5	12.5	50	10	US-09-895-814-585	Sequence 585, App
893	27	12.7	77	9	US-10-116-252-21	Sequence 21, Appl	966	26.5	12.5	50	10	US-09-759-143-585	Sequence 585, App
894	27	12.7	78	9	US-09-925-299-1065	Sequence 1065, App	967	26.5	12.5	50	10	US-09-864-761-45083	Sequence 45083, A
895	27	12.7	78	9			968	26.5	12.5	50	10	US-09-780-669-585	Sequence 585, App

```

969 26.5 12.5 50 10 US-09-822-827-585
970 26.5 12.5 52 10 US-09-864-761-39691
971 26.5 12.5 52 10 US-09-864-761-49064
972 26.5 12.5 53 10 US-10-083-357-1224
973 26.5 12.5 53 10 US-09-864-761-39811
974 26.5 12.5 53 10 US-09-864-761-48090
975 26.5 12.5 54 10 US-09-864-761-45795
976 26.5 12.5 54 10 US-09-764-877-1205
977 26.5 12.5 56 10 US-09-864-761-33736
978 26.5 12.5 56 10 US-09-864-761-35434
979 26.5 12.5 56 10 US-09-864-761-39524
980 26.5 12.5 57 9 US-09-738-626-5498
981 26.5 12.5 57 10 US-09-864-761-37163
982 26.5 12.5 57 10 US-09-864-761-37572
983 26.5 12.5 57 10 US-09-864-761-43269
984 26.5 12.5 59 9 US-09-798-692-1016
985 26.5 12.5 59 9 US-09-798-692-1110
986 26.5 12.5 59 9 US-09-798-692-1123
987 26.5 12.5 59 9 US-09-798-692-1123
988 26.5 12.5 59 10 US-09-764-877-1229
989 26.5 12.5 60 9 US-10-012-896-586
990 26.5 12.5 60 9 US-10-012-896-886
991 26.5 12.5 60 9 US-10-012-896-892
992 26.5 12.5 60 9 US-10-013-379-22
993 26.5 12.5 60 9 US-09-895-793-586
994 26.5 12.5 60 9 US-09-895-793-886
995 26.5 12.5 60 9 US-09-895-793-886
996 26.5 12.5 60 9 US-09-895-814-586
997 26.5 12.5 60 9 US-09-895-814-586
998 26.5 12.5 60 9 US-09-895-814-892
999 26.5 12.5 60 9 US-09-809-391-671
1000 26.5 12.5 60 10 US-09-759-143-586

```

## ALIGNMENTS

```

RESULT 1
; Sequence 47521, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30

```

```

; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
; SEQ ID NO 47521
; LENGTH: 84
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL158153.2
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 6.7
; OTHER INFORMATION: EST_HUMAN HIT: BF573955.1, EVALUE 1.60e-02
; OTHER INFORMATION: SWISSPROT HIT: Q91641, EVALUE 3.00e-25
US-09-864-761-47521
Query Match 21.2%; Score 45; DB 10; Length 84;
Best Local Similarity 26.4%; Pred. No. 24;
Matches 14; Conservative 6; Mismatches 7; Indels 26; Gaps 2;
QY 17 GQKSRVINEP-----TEALSVAV-----EGLAWRK 43
DB 11 GQKARLLSRPLRGVSKHCLTFYHMYGGTGLLSVYLKKEESESLLWRR 63
RESULT 2
US-09-879-957-111
; Sequence 111, Application US/09879957
; Patent No. US20020034755A1
; GENERAL INFORMATION:
; APPLICANT: SPARKS, Andrew B.
; APPLICANT: HOFFMAN, No. US20020034755a1h
; APPLICANT: KAY, Brian K.
; APPLICANT: FOWLKES, Dana M.
; APPLICANT: MCCONNELL, Stephen J.
; TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL
; DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND
; USING SAME
; NUMBER OF SEQUENCES: 227
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/879,957
; FILING DATE: 13-Jun-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/630,915
; FILING DATE: 03-APR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Mistock, S. Leslie
; REGISTRATION NUMBER: 18,872

```

```
; REFERENCE/DOCKET NUMBER: 1101-174
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 798-9099
; TELEFAX: (212) 869-8864/9741
; TELEX: 86141 PENNIE
; INFORMATION FOR SEQ ID NO: 111:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 55 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 111:
US-09-879-957-111
Query Match 20.5% Score 43.5; DB 10; Length 55;
Best Local Similarity 41.4%; Pred. No. 23;
Matches 12; Conservative 8; Mismatches 4; Indels 5; Gaps 2;
Qy 4 SISENSLVAMDFS-COKSRVIENTEALS 31
Db 23 IVNKGSLVAGFGDQGEAR-----PEEILN 47

RESULT 3
US-09-864-761-34262
; Sequence 34262, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecmiga-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
```

```
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 34262
; LENGTH: 63
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL031734.9
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 0.87
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 74
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 0.84
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.82
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.9
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.89
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.8
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.79
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.1
; OTHER INFORMATION: EST_HUMAN HIT: A1075970.1, EVALUJE 7.00e-03
US-09-864-761-34262
Query Match 19.8%; Score 42; DB 10; Length 63;
Best Local Similarity 31.8%; Pred. No. 44;
Matches 14; Conservative 7; Mismatches 11; Indels 12; Gaps 2;
Qy 6 SENSELVAMDFS-GOKSRVI-----ENPTEALSVAVEEGLAWRK 42
Db 13 SQVGLPILYFSGRRRLRLRPEVLAEIPREAFVTE-----AWVK 51

RESULT 4
US-09-738-626-6764
; Sequence 6764, Application US/09738626
; Publication No. US20020197603A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENO, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 6764
; LENGTH: 72
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-6764
Query Match 19.8%; Score 42; DB 9; Length 72;
Best Local Similarity 22.0%; Pred. No. 53;
Matches 11; Conservative 12; Mismatches 19; Indels 8; Gaps 1;
Qy 2 MRSISENSLVAMDFS-GOKSRVI-----ENPTEALSVAVEEGLAWRK 43
Db 1 MHFIKENLITSAESNALRAQLMLSLGSAEFERSIIRERQAEGLAWRK 50
```

Query Match 18.9%; Score 40; DB 10; Length 49;  
Best Local Similarity 38.7%;  
Matches 12; Conservative 7; Mismatches 8; Indels

```

; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/055,954
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-08-18
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/058,785
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-09-12
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/058,664
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-09-12
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/058,660
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-09-12
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/058,661
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-09-12
; NUMBER OF SEQ ID NOS: 672
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 611
; LENGTH: 63
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-983-802-611

```

```

Query Match      18.9%; Score 40; DB 9; Length 63;
Best Local Similarity 47.18; Pred. No. 86;
Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

```

```

QY 5 ISNSLVAMDFSGQSKR 21
DB 20 VSSNLSFPFFGOKAR 36

```

```

RESULT 7
US-09-895-913A-280
; Sequence 280, Application US/09895913A
; Patent No. US20020160456A1
; GENERAL INFORMATION:
; APPLICANT: Kleanthous, Harold
; APPLICANT: Al-Garawi, Amal
; APPLICANT: Miller, Charles
; APPLICANT: Tomb, Jean Francois
; APPLICANT: Omen, Raymond P.
; TITLE OF INVENTION: Identification of Polynucleotides
; TITLE OF INVENTION: Encoding No. US20020160456A1 Helicobacter Polypeptides in the
; FILE REFERENCE: 06132/043002
; CURRENT APPLICATION NUMBER: US/09/895,913A
; CURRENT FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 08/861,227
; PRIOR FILING DATE: 1997-06-24
; NUMBER OF SEQ ID NOS: 368
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 280
; LENGTH: 70
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-09-895-913A-280

```

```

Query Match      18.9%; Score 40; DB 9; Length 70;
Best Local Similarity 25.08; Pred. No. 99;
Matches 8; Conservative 8; Mismatches 16; Indels 0; Gaps 0;

```

```

QY 1 PMRISISLVAMDFSGQSKRSVIENTPELSV 32
DB 19 PMFLPKNEWLVKFSNSDALPINKPSTLSI 50

```

```

RESULT 8
US-09-764-872-351
; Sequence 351, Application US/09764872
; Publication No. US20030050231A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA125
; CURRENT APPLICATION NUMBER: US/09/764,872
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper

```

```

; NUMBER OF SEQ ID NOS: 957
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 351
; LENGTH: 71
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (50)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (52)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (58)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-872-351

```

```

Query Match      18.9%; Score 40; DB 9; Length 71;
Best Local Similarity 44.0%; Pred. No. 1e+02;
Matches 11; Conservative 3; Mismatches 5; Indels 6; Gaps 1;

```

```

QY 19 KSRVIENTPELSVAVEEGLAWRKK 43
DB 20 KVRVKNFREAVN-----FHWRRK 38

```

```

RESULT 9
US-09-864-761-47104
; Sequence 47104, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL I
; FILE REFERENCE: Aesonic-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21

```

```

; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 47104
; LENGTH: 68
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC005083.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.7
; OTHER INFORMATION: EXPRESSED IN LONG, SIGNAL = 0.49
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.5
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.48
; OTHER INFORMATION: SWISSPROT HIT: P43802, EVALU8 8.80e-01
US-09-864-761-47104

Query Match
Best Local Similarity 18.6%; Score 39.5; DB 10; Length 68;
Matches 11; Conservative 8; Mismatches 14; Indels 1; Gaps 1;

Qy 9 SLVAMDFSGQSRVIENTEALSVAVEGLWRK 42
Db 1 ALIKMDCQELVARLQIE-AAVLTSVAVKLGKGWRE 33

RESULT 10
US-09-764-887-218
; Sequence 218, Application US/09764887
; Patent No. US20020042096A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; FILE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; CURRENT APPLICATION NUMBER: US/09/764,887
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 658
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 218
; LENGTH: 76
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-887-218

Query Match
Best Local Similarity 18.6%; Score 39.5; DB 10; Length 78;
Matches 12; Conservative 7; Mismatches 9; Indels 5; Gaps 1;

Qy 3 RSISENSLVAMDFSGQSRVIENTEALSVAVE 35
Db 8 RSDSDSSTLA-----KKSIFVNSTERSLRV 35

RESULT 11
US-09-815-242-4918
; Sequence 4918, Application US/09815242
; Patent No. US20020081569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE OF INVENTION: Prokaryotes
; CURRENT APPLICATION NUMBER: US/09/815,242

```

```

; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4918
; LENGTH: 89
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-815-242-4918

Query Match
Best Local Similarity 18.6%; Score 39.5; DB 10; Length 89;
Matches 11; Conservative 7; Mismatches 13; Indels 3; Gaps 1;

Qy 5 ISENSLVAMDFSGQSRVIENTEALSVAVEGL 38
Db 8 ITEKSLAND---EKKYTFEVDTRANKTLVKQAV 38

RESULT 12
US-09-864-761-43815
; Sequence 43815, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; FILE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/006666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006661

```



;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00670  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: US 60/234,687  
;; PRIOR FILING DATE: 2000-09-21  
;; PRIOR APPLICATION NUMBER: US 09/608,408  
;; PRIOR FILING DATE: 2000-06-30  
;; PRIOR APPLICATION NUMBER: US 09/774,203  
;; PRIOR FILING DATE: 2001-01-29  
;; NUMBER OF SEQ ID NOS: 49117  
;; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1  
;; SEQ ID NO 43815  
;; LENGTH: 47  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
;; FEATURE:  
;; OTHER INFORMATION: MAP TO AC005881.2  
;; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.73  
;; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.6  
;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.9  
;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.68  
;; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.76  
;; OTHER INFORMATION: SWISSPROT HIT: Q60349, EVALUATE 4.70e+00  
US-09-864-761-43815

Query Match 18.4%; Score 39; DB 10; Length 47;  
Best Local Similarity 34.8%; Pred. No. 83;  
Matches 8; Conservative 3; Mismatches 12; Indels 0; Gaps 0;

QY 10 LVAMDFSGQKSRVIENTEALVS 32  
I: | | | | | | | | | |  
Db 3 LLMFSGHGVSVKTEKPTVKLMI 25

## RESULT 13

US-09-864-761-45291  
; Sequence 45291, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharon G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; FILE REFERENCE: Aecomica-X-1  
; CURRENT APPLICATION NUMBER: PCT/US01/00667  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662

;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00661  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: PCT/US01/00670  
;; PRIOR FILING DATE: 2001-01-30  
;; PRIOR APPLICATION NUMBER: US 60/234,687  
;; PRIOR FILING DATE: 2000-09-21  
;; PRIOR APPLICATION NUMBER: US 09/608,408  
;; PRIOR FILING DATE: 2000-06-30  
;; PRIOR APPLICATION NUMBER: US 09/774,203  
;; PRIOR FILING DATE: 2001-01-29  
;; NUMBER OF SEQ ID NOS: 49117  
;; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1  
;; SEQ ID NO 45291  
;; LENGTH: 60  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
;; FEATURE:  
;; OTHER INFORMATION: MAP TO AC009079.4  
;; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.54  
;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.77  
;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.65  
;; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.74  
;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.48  
;; OTHER INFORMATION: EST\_HUMAN HIT: AV756022.1, EVALUATE 6.00e-28  
;; OTHER INFORMATION: SWISSPROT HIT: P56315, EVALUATE 2.00e+00  
US-09-864-761-45291

Query Match 18.4%; Score 39; DB 10; Length 60;  
Best Local Similarity 33.3%; Pred. No. 1.1e+02;  
Matches 7; Conservative 7; Mismatches 7; Indels 0; Gaps 0;

QY 5 ISESLVAMDFSGQKSRVNIEN 25  
| | | | | | | | | |  
Db 34 LSENLQTYNFRQKESVVOH 54

## RESULT 14

US-09-864-761-36073  
; Sequence 36073, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharon G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL F  
; FILE REFERENCE: Aecomica-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30

## POLYPEPTIDES and Their Uses

ON SEQUENCES: 534  
CORRESPONDENCE ADDRESS:  
Beecham Corporation  
Attention: Marketing  
Street: 709 Swedeland  
City: King of Prussia  
State: PA  
Country: USA  
ZIP: 19406-0939  
COMPUTER READABLE FORM:

COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/939,980  
FILING DATE: 27-Aug-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/936,165  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Glimmi, Edward R  
REGISTRATION NUMBER: 38,891  
REFERENCE/DOCKET NUMBER: P50549  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 610-270-4478  
TELEFAX: 610-270-5090  
TELEX: <Unknown>  
INFORMATION FOR SEQ ID NO: 301:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 62 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: Protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 301:  
5-09-939-980-301

```

Query Match      17.9%; Score 38; DB 10; Length 62;
Best Local Similarity 31.0%; Pred. No. 1.6e+02;
Matches 9; Conservative 6; Mismatches 14; Indels 0; Gaps 0;

/ 10 LVAMDFSGQKSRVIENTEALSAVVEGL 38
      : : : : : | : | : | : |
7 MMALNTAGXTQRTVETLAEYSGVPVYNGL 35

```

; Sequence: 1560, Application US/09925300  
; Patent No. US20020151681A1  
; GENERAL INFORMATION:  
; APPLICANT: Craig Rosen,  
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
; FILE REFERENCE: PA101  
; CURRENT APPLICATION NUMBER: US/09/925,300  
; PRIOR FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: PCT/US00/05988  
; PRIOR FILING DATE: 2000-03-08  
; PRIOR FILING DATE: 2000-03-08  
; PRIOR FILING DATE: 1999-03-12  
; NUMBER OF SEQ ID NOS: 1890  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 1560  
; LENGTH: 68  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-925-300-1560

Query Match 17.9%; Score 38; DB 10; Length 68;  
Best Local Similarity 34.6%; Pred. No. 1.9e+02;  
Matches 9; Conservative 5; Mismatches 12; Indels 0; Gaps 0;

QY 18 OKSRVIENTEALSVAVEEGLAWRK 43  
DB 31 KKNSSVKNLCCGLSIFAAFGLRWRK 56

RESULT 18  
US-09-864-761-43411  
; Sequence 43411, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharon G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; FILE REFERENCE: Aecmica-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR FILING DATE: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR FILING DATE: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR FILING DATE: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR FILING DATE: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR FILING DATE: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR FILING DATE: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR FILING DATE: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR FILING DATE: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR FILING DATE: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR FILING DATE: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR FILING DATE: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR FILING DATE: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR FILING DATE: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR FILING DATE: PCT/US01/00661  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR FILING DATE: PCT/US01/00670

; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: US 09/608,408  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: US 09/774,203  
; PRIOR FILING DATE: 2001-01-29  
; NUMBER OF SEQ ID NOS: 49117  
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 43411  
; LENGTH: 51  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO AC006257.1  
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.74  
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.77  
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.76  
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.85  
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.83  
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.71  
; OTHER INFORMATION: SWISSPROT HIT: P45643, EVALUATION 1.00e-21  
US-09-864-761-43411

Query Match 17.7%; Score 37.5; DB 10; Length 51;  
Best Local Similarity 26.1%; Pred. No. 1.5e+02;  
Matches 12; Conservative 7; Mismatches 8; Indels 19; Gaps 1;

QY 13 MDSGQKSR-----VIENTEALSVAVEEGLA 39  
DB 3 MEFSGRKWRKRLAGDQRNARYPHCLQFYLPSPENISLIEFENLA 48

RESULT 19  
US-09-925-300-1765  
; Sequence 1765, Application US/09925300  
; Patent No. US20020151681A1  
; GENERAL INFORMATION:  
; APPLICANT: Craig Rosen,  
; APPLICANT: Steve Ruben  
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
; FILE REFERENCE: PA101  
; CURRENT APPLICATION NUMBER: US/09/925,300  
; CURRENT FILING DATE: 2001-08-10  
; PRIOR FILING DATE: PCT/US00/05988  
; PRIOR FILING DATE: 2000-03-08  
; PRIOR APPLICATION NUMBER: 60/124,270  
; PRIOR FILING DATE: 1999-03-12  
; NUMBER OF SEQ ID NOS: 1890  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 1765  
; LENGTH: 64  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; NAME/KEY: SITE  
; LOCATION: (26)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-925-300-1765

Query Match 17.7%; Score 37.5; DB 10; Length 64;  
Best Local Similarity 60.0%; Pred. No. 2e+02;  
Matches 9; Conservative 3; Mismatches 2; Indels 1; Gaps 1;

QY 29 ALSV-AVEEGLAWRK 42  
DB 11 ALSIDSVEDHLAWSK 25

RESULT 20  
US-10-092-154-735  
; Sequence 735, Application US/10092154  
; Publication No. US20030054375A1

```

; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC009C1
; CURRENT APPLICATION NUMBER: US/10/092,154
; CURRENT FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 2003
; Prior Application removed - See File Wrapper or Palm
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 735
; LENGTH: 65
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-092-154-735

```

```

Query Match      17.7%; Score 37.5; DB 9; Length 65;
Best Local Similarity 30.0%; Pred. No. 2.le+02;
Matches 15; Conservative 6; Mismatches 10; Indels 19; Gaps 2;

Qy 12 AMDFS-GOKSRVIENTPE-----ALSVAVEGLAWRK 42
Db 7 AMKFSKNQYFVLELPSEDVKLVNIKTVOITNHFPLPVATKSSLSFQK 56

```

```

RESULT 21
US-09-764-847-735
; Sequence 735, Application US/09764847
; Patent No. US20020132767A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC009
; CURRENT APPLICATION NUMBER: US/09/764,847
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 2003
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 735
; LENGTH: 65
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-847-735

```

```

Query Match      17.7%; Score 37.5; DB 10; Length 65;
Best Local Similarity 30.0%; Pred. No. 2.le+02;
Matches 15; Conservative 6; Mismatches 10; Indels 19; Gaps 2;

Qy 12 AMDFS-GOKSRVIENTPE-----ALSVAVEGLAWRK 42
Db 7 AMKFSKNQYFVLELPSEDVKLVNIKTVOITNHFPLPVATKSSLSFQK 56

```

```

RESULT 22
US-09-925-297-543
; Sequence 543, Application US/09925297
; Patent No. US20020081659A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA105
; CURRENT APPLICATION NUMBER: US/09/925,297
; CURRENT FILING DATE: 2001-08-10
; Prior Application Number: PCT/US00/05989
; Prior Filing Date: 2000-03-08
; Prior Application Number: 60/124,270
; Prior Filing Date: 1999-03-12
; NUMBER OF SEQ ID NOS: 928
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 543
; LENGTH: 89
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-925-297-543

```

```

Query Match      17.7%; Score 37.5; DB 10; Length 89;
Best Local Similarity 27.0%; Pred. No. 3.le+02;
Matches 10; Conservative 9; Mismatches 15; Indels 3; Gaps 1;

Qy 3 RSISNSLVAMDFSGOKSRVIENTPEALSVAVEGLA 39
Db 22 KQAKENPAYRLELNGHRRRL---TWEATPRSIHEGIA 55

```

```

RESULT 23
US-09-071-838-294
; Sequence 294, Application US/09071838
; Patent No. US20020152501A1
; GENERAL INFORMATION:
; APPLICANT: Fischer, Robert L.
; APPLICANT: Ohad, Nir
; APPLICANT: Kiyosue, Tomohiro
; APPLICANT: Yadegari, Ramin
; APPLICANT: Margossian, Linda
; APPLICANT: Harada, John
; APPLICANT: Goldberg, Robert B.
; TITLE OF INVENTION: Nucleic Acids That Control Seed and
; NUMBER OF SEQUENCES: 324
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/071,838
; FILING DATE: 01-MAY-1998
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Bastian, Kevin L.
; REGISTRATION NUMBER: 34,774
; REFERENCE/DOCKET NUMBER: 023070-0861000S
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 294:
; LENGTH: 57 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-071-838-294

```

```

Query Match      17.5%; Score 37; DB 10; Length 57;
Best Local Similarity 45.8%; Pred. No. 2.le+02;
Matches 11; Conservative 4; Mismatches 7; Indels 2; Gaps 1;

Qy 2 MRISNSLVAMDFSGOKSRVIENT 25
Db 17 VRSFS--STAMDQNRIGRVVEN 38

```

```

RESULT 24
US-09-764-877-1480
; Sequence 1480, Application US/09764877
; Patent No. US20020147140A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC005

```

```

; CURRENT APPLICATION NUMBER: US/09/764,877
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 4031
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1480
; LENGTH: 86
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-877-1480

```

```

Query Match 17.5% Score 37; DB 10; Length 86;
Best Local Similarity 32.0% Pred. No. 3.5e+02;
Matches 8; Conservative 5; Mismatches 8; Indels 4; Gaps 1;

```

```

Qy 16 SGOKSRVIENPTEALSVAVEEGLAW 40
   ||| : | : : : ||||
Db 18 SQO----VFGPRRSQTLLFQSGLAW 38

```

```

RESULT 25
US-09-867-550-980
; Sequence 980, Application US/09867550
; Patent No. US20020082206A1
; GENERAL INFORMATION:
; APPLICANT: Leach, Martin D.
; APPLICANT: Mehraban, Fuad,
; APPLICANT: Conley, Pamela
; APPLICANT: Law, Debbie
; APPLICANT: Topper, James
; TITLE OF INVENTION: No. US20020082206A1el Polynucleotides from Atherogenic Cells and
; TITLE OF INVENTION: Thereby
; FILE REFERENCE: 21402-013 (Cura-313)
; CURRENT APPLICATION NUMBER: US/09/867,550
; CURRENT FILING DATE: 2001-09-20
; PRIOR APPLICATION NUMBER: USSN 60/208,427
; PRIOR FILING DATE: 2000-05-30
; NUMBER OF SEQ ID NOS: 2125
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 980
; LENGTH: 59
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-867-550-980

```

```

Query Match 17.2% Score 36.5; DB 10; Length 59;
Best Local Similarity 27.6% Pred. No. 2.6e+02;
Matches 8; Conservative 6; Mismatches 4; Indels 11; Gaps 1;

```

```

Qy 15 FSGOKSRVIENPTEALSVAVEEGLAWBK 43
   ||||| : : |||||
Db 32 YSGKKPR-----PIAGLSWQOK 49

```

```

Search completed: March 28, 2003, 09:19:16
Job time : 22.8619 secs

```

















777 26 12.3 50 3 US-09-314-093-44  
 778 26 12.3 50 4 US-09-250-848-44  
 779 26 12.3 50 4 US-09-004-406C-22  
 780 26 12.3 50 4 US-09-251-885-44  
 781 26 12.3 50 4 US-09-337-635-44  
 782 26 12.3 50 4 US-09-337-635-44  
 783 26 12.3 50 4 US-09-469-260A-534  
 784 26 12.3 50 5 PCT-US91-02942-8  
 785 26 12.3 51 1 US-08-056-200-113  
 786 26 12.3 51 1 US-08-570-137-11  
 787 26 12.3 51 2 US-08-736-334B-9  
 788 26 12.3 51 2 US-08-800-644-113  
 789 26 12.3 51 4 US-09-127-048-2  
 790 26 12.3 51 4 US-09-076-510-11  
 791 26 12.3 52 3 US-08-917-299-24  
 792 26 12.3 52 4 US-09-188-930-177  
 793 26 12.3 52 4 US-09-422-662-24  
 794 26 12.3 52 4 US-09-330-330-13  
 795 26 12.3 53 2 US-08-942-423-33  
 796 26 12.3 53 4 US-08-965-056-94  
 797 26 12.3 53 4 US-09-346-510B-20  
 798 26 12.3 53 4 US-08-858-207A-444  
 799 26 12.3 54 2 US-08-456-647B-54  
 1000 26 12.3 54 2 US-08-237-401A-54

ALIGNMENTS

RESULT 1  
 US-09-331-930A-22  
 ; Sequence 22, Application US/09331930A  
 ; Patent No. 6436670  
 ; GENERAL INFORMATION:  
 ; APPLICANT: COLLIER, PAUL Z.  
 ; TITLE OF INVENTION: A NOVEL GENE AND USES THEREFOR  
 ; FILE REFERENCE: 22975-20007.00  
 ; CURRENT APPLICATION NUMBER: US/09/331,930A  
 ; CURRENT FILING DATE: 1999-06-30  
 ; PRIOR APPLICATION NUMBER: PCT/AU98/00902  
 ; PRIOR FILING DATE: 1998-10-30  
 ; PRIOR APPLICATION NUMBER: AU PP0117/97  
 ; PRIOR FILING DATE: 1997-10-31  
 ; PRIOR APPLICATION NUMBER: AU PP0323/97  
 ; PRIOR FILING DATE: 1997-11-11  
 ; NUMBER OF SEQ ID NOS: 27  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 22  
 ; LENGTH: 73  
 ; TYPE: PRT  
 ; ORGANISM: Caenorhabditis elegans  
 US-09-331-930A-22

Query Match 21.5%; Score 45.5; DB 4; Length 73;  
 Best Local Similarity 29.4%; Pred. No. 8.6;  
 Matches 10; Conservative 7; Mismatches 12; Indels 5; Gaps 1;  
 QY 14 DFGSGKSRVIENTEALS-----VAVEGLAWRK 42  
 DB 8 DRLGKVKRIKNSDTIGDLKLIAAQTGRWEK 41

RESULT 2  
 US-08-776-059-18  
 ; Sequence 18, Application US/08776059B  
 ; Patent No. 6271368  
 ; GENERAL INFORMATION:  
 ; APPLICANT: LENTZEN, Hans  
 ; APPLICANT: ECK, Jurgen  
 ; APPLICANT: BAUR, Axel  
 ; APPLICANT: ZINKE, Holger  
 ; TITLE OF INVENTION: RECOMBINANT MISTLETOE LECTIN (RML)

FILE REFERENCE: 674503-2003  
 ; CURRENT APPLICATION NUMBER: US/08/776,059B  
 ; CURRENT FILING DATE: 1999-06-19  
 ; EARLIER APPLICATION NUMBER: PCT/EP96/02273  
 ; EARLIER FILING DATE: 1996-06-25  
 ; EARLIER APPLICATION NUMBER: 95109949.8  
 ; EARLIER FILING DATE: 1995-06-26  
 ; NUMBER OF SEQ ID NOS: 56  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 18  
 ; LENGTH: 47  
 ; TYPE: PRT  
 ; ORGANISM: Saponaria officinalis  
 US-08-776-059-18  
 Query Match 20.8%; Score 44; DB 4; Length 47;  
 Best Local Similarity 34.8%; Pred. No. 8;  
 Matches 8; Conservative 7; Mismatches 8; Indels 0; Gaps 0;  
 QY 13 MDFSOGKSRVIENTEALSVAVE 35  
 DB 6 MDVNVKKARVVKNEARELLIAIQ 28

RESULT 3  
 US-09-081-320-20  
 ; Sequence 20, Application US/09081320  
 ; Patent No. 6093544  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Gonsalves, Dennis  
 ; APPLICANT: Meng, Baozhong  
 ; TITLE OF INVENTION: RUPESTRIS STEM PITTING ASSOCIATED VIRUS  
 ; TITLE OF INVENTION: NUCLEIC ACIDS, PROTEINS, AND THEIR USES  
 ; NUMBER OF SEQUENCES: 54  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP  
 ; STREET: Clinton Square, P.O. Box 1051  
 ; CITY: Rochester  
 ; STATE: New York  
 ; COUNTRY: U.S.A.  
 ; ZIP: 14603  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/081,320  
 ; FILING DATE:  
 ; CLASSIFICATION:  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 60/047,147  
 ; FILING DATE: 20-MAY-1997  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 60/069,902  
 ; FILING DATE: 17-DEC-1997  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Goldman, Michael L.  
 ; REGISTRATION NUMBER: 30,727  
 ; REFERENCE/DOCKET NUMBER: 19603/1722  
 ; TELEPHONE: (716) 263-1304  
 ; TELEFAX: (716) 263-1600  
 ; INFORMATION FOR SEQ ID NO: 20:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 80 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS:  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 US-09-081-320-20  
 Query Match 20.8%; Score 44; DB 3; Length 80;

Query Match 20.8%; Score 44; DB 4; Length 80;  
Best Local Similarity 50.0%; Pred. No. 17;  
Matches 13; Conservative 4; Mismatches 7; Indels

RESULT 7  
US-08-776-059-16 ; Sequence 16, Application US/08776059B  
; Patent No. 6271368  
; GENERAL INFORMATION:  
; APPLICANT: LENTZEN, Hans  
; APPLICANT: ECK, Jurgen  
; APPLICANT: BAUR, Axel  
; APPLICANT: ZINKER, Holger  
; TITLE OF INVENTION: RECOMBINANT MISTLETOE LECTIN (RML)  
; FILE REFERENCE: 674503-2003  
; CURRENT APPLICATION NUMBER: US/08/776.059B

Fri Mar 28 12:13:32 2003

us-09-936-697-5.max.rai

Page 10

; CURRENT FILING DATE: 1999-06-19  
; EARLIER APPLICATION NUMBER: PCT/EP96/02273  
; EARLIER FILING DATE: 1996-06-25  
; EARLIER APPLICATION NUMBER: 95109949.8  
; NUMBER OF SEQ ID NOS: 26  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 16  
; LENGTH: 47  
; TYPE: PRT  
; ORGANISM: Saponaria officinalis  
US-08-776-059-16

Query Match 19.8%; Score 42; DB 4; Length 47;  
Best Local Similarity 30.8%; Pred. No. 16;  
Matches 8; Conservative 10; Indels 0; Gaps 0;  
QY 10 LVAMDFSGOKSRVIENTPALSVAVE 35  
DB 3 LTFMEAVNKARVVKNFRLIIAQ 28

RESULT 8  
US-07-641-971B-5  
; Sequence 5, Application US/07641971B  
; Patent No. 5236706  
; GENERAL INFORMATION:  
; APPLICANT: Debre, Patrice  
; APPLICANT: Mossalayi, Mohammed D  
; TITLE OF INVENTION: A PHARMACEUTICAL PREPARATION FOR THE  
; TITLE OF INVENTION: MATURATION OF PROTHYMOCYTES  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Irving M. Fishman, CIBA-GEIGY Corporation  
; STREET: 556 Morris Avenue  
; CITY: Summit  
; STATE: New Jersey  
; COUNTRY: USA  
; ZIP: 07901  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/641,971B  
; FILING DATE: 19910116  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 90016254  
; FILING DATE: 24-JAN-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Fishman, Irving M  
; REGISTRATION NUMBER: 30258  
; REFERENCE/DOCKET NUMBER: 4-17921/+/DEB  
; TELEPHONE: 908-277-4832  
; TELEFAX: 908-277-4306  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 40 amino acids  
; TYPE: AMINO ACID  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; FRAGMENT TYPE: N-terminal  
US-07-641-971B-5

Query Match 19.3%; Score 41; DB 1; Length 40;  
Best Local Similarity 39.3%; Pred. No. 19;  
Matches 11; Conservative 5; Mismatches 8; Indels 4; Gaps 1;

QY 1 PMRSISENSLVAMDFSGOKSRVIENTPTE 28  
DB 2 PVRSNLN---CTLRDSGOKSLVMSGPYE 25  
RESULT 9  
US-07-781-248A-5  
; Sequence 5, Application US/07781248A  
; Patent No. 5246699  
; GENERAL INFORMATION:  
; APPLICANT: Debre, Patrice  
; APPLICANT: Mossalayi, Mohammed D  
; TITLE OF INVENTION: MATURATION OF HEMATOPOIETIC CELLS  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Irving M. Fishman, CIBA-GEIGY Corporation  
; STREET: 556 Morris Avenue  
; CITY: Summit  
; STATE: New Jersey  
; COUNTRY: USA  
; ZIP: 07901  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/781,248A  
; FILING DATE: 19911230  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 90103565  
; FILING DATE: 09-MAY-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ikeiger, Barbara J.  
; REGISTRATION NUMBER: 36,170  
; REFERENCE/DOCKET NUMBER: 4-18065/A/DEB  
; TELEPHONE: 908-277-3368  
; TELEFAX: 908-277-4306  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 40 amino acids  
; TYPE: AMINO ACID  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; FRAGMENT TYPE: N-terminal  
US-07-781-248A-5

Query Match 19.3%; Score 41; DB 1; Length 40;  
Best Local Similarity 39.3%; Pred. No. 19;  
Matches 11; Conservative 5; Mismatches 8; Indels 4; Gaps 1;  
QY 1 PMRSISENSLVAMDFSGOKSRVIENTPTE 28  
DB 2 PVRSNLN---CTLRDSGOKSLVMSGPYE 25

RESULT 10  
US-08-459-568-52  
; Sequence 52, Application US/08459568  
; Patent No. 5811304  
; GENERAL INFORMATION:  
; APPLICANT: Huang, Shi  
; TITLE OF INVENTION: Retinoblastoma Protein - Interacting  
; TITLE OF INVENTION: Zinc Finger Proteins  
; NUMBER OF SEQUENCES: 93  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Campbell and Flores

STREET: 4370 La Jolla Village Drive, Suite 700  
CITY: San Diego  
STATE: California  
COUNTRY: USA  
ZIP: 92122  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
FILING DATE: 02-JUN-1995  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA: US 08/399,411  
FILING DATE: 06-MAR-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Campbell, Cathryn A.  
REGISTRATION NUMBER: 31,815  
REFERENCE/DOCKET NUMBER: P-LJ 1264  
TELEPHONE: (619) 535-9001  
TELEFAX: (619) 535-8949  
INFORMATION FOR SEQ ID NO: 52:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 66 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
US-08-459-568-52

Query Match 19.1%; Score 40.5; DB 2; Length 66;  
Best Local Similarity 47.4%; Pred. No. 44;  
Matches 9; Conservative 7; Mismatches 2; Indels 1; Gaps 1;

QY 4 SISENSLVAMDFS-GOKSR 21  
DB 22 TVNKGSLVALGFSGQEAR 40

RESULT 11  
US-08-399-411-52  
Sequence 52, Application US/08399411  
Patent No. 5831008  
GENERAL INFORMATION:  
APPLICANT: Huang, Shi  
TITLE OF INVENTION: Retinoblastoma protein - Interacting  
NUMBER OF SEQUENCES: 93  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Campbell & Flores  
STREET: 4370 La Jolla Village Drive, Suite 700  
CITY: San Diego  
STATE: California  
COUNTRY: USA  
ZIP: 92122  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
FILING DATE: 06-MAR-1995  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Campbell, Cathryn A.  
REGISTRATION NUMBER: 31,815  
REFERENCE/DOCKET NUMBER: P-LJ 1264  
TELEPHONE: (619) 535-9001  
TELEFAX: (619) 535-8949  
INFORMATION FOR SEQ ID NO: 52:

SEQUENCE CHARACTERISTICS:  
LENGTH: 66 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
US-08-399-411-52

Query Match 19.1%; Score 40.5; DB 2; Length 66;  
Best Local Similarity 47.4%; Pred. No. 44;  
Matches 9; Conservative 7; Mismatches 2; Indels 1; Gaps 1;

QY 4 SISENSLVAMDFS-GOKSR 21  
DB 22 TVNKGSLVALGFSGQEAR 40

RESULT 12  
US-08-516-859A-52  
Sequence 52, Application US/08516859A  
Patent No. 6069231  
GENERAL INFORMATION:  
APPLICANT: Huang, Shi  
TITLE OF INVENTION: Retinoblastoma Protein - Interacting  
NUMBER OF SEQUENCES: 106  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Campbell & Flores LLP  
STREET: 4370 La Jolla Village Drive, Suite 700  
CITY: San Diego  
STATE: California  
COUNTRY: USA  
ZIP: 92122  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
FILING DATE: 18-AUG-1995  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/399,411  
FILING DATE: 06-MAR-1995  
APPLICATION DATA: US 08/292,683  
FILING DATE: 18-AUG-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Campbell, Cathryn A.  
REGISTRATION NUMBER: 31,815  
REFERENCE/DOCKET NUMBER: P-LJ 1776  
TELEPHONE: (619) 535-9001  
TELEFAX: (619) 535-8949  
INFORMATION FOR SEQ ID NO: 52:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 66 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
US-08-516-859A-52

Query Match 19.1%; Score 40.5; DB 3; Length 66;  
Best Local Similarity 47.4%; Pred. No. 44;  
Matches 9; Conservative 7; Mismatches 2; Indels 1; Gaps 1;

QY 4 SISENSLVAMDFS-GOKSR 21  
DB 22 TVNKGSLVALGFSGQEAR 40

RESULT 13  
US-09-586-472-52  
Sequence 52, Application US/09586472  
Patent No. 6323335



ZIP: 92122  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible

APPLICANT: Lacroix, Bruno  
TITLE OF INVENTION: 5' ESTS FOR SECRETED PROTEINS  
NUMBER OF SEQUENCES: 503  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Knobbe, Martens, Olson & Bear  
STREET: 501 West Broadway  
CITY: San Diego  
STATE: California  
COUNTRY: USA  
ZIP: 92101-3505  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy Disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: Win95  
SOFTWARE: Word  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/905,223  
FILING DATE:  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: Israel, Ned A.  
REGISTRATION NUMBER: 29,655  
REFERENCE/DOCKET NUMBER:  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 235-8550  
TELEFAX: (619) 235-0176  
INFORMATION FOR SEQ ID NO: 420:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 86 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: LINEAR  
MOLECULE TYPE: PROTEIN  
ORIGINAL SOURCE:  
ORGANISM: Homo Sapiens  
TISSUE TYPE: Brain  
FEATURE:  
NAME/KEY: sig\_peptide  
LOCATION: -26...-1  
IDENTIFICATION METHOD: Von Heijne matrix  
OTHER INFORMATION: score 4.5  
OTHER INFORMATION: seq ENSLILLLQGLQG/RV  
US-08-905-223-420

Query Match 19.1%; Score 40.5; DB 4; Length 86;  
Best Local Similarity 50.0%; Pred. No. 64;  
Matches 10; Conservative 4; Mismatches 5; Indels 1; Gaps 1;

QY 3 RSIENSLVAMDFSGQKSRV 22  
Db 10 RTIENSLILLL-LQGLQGRV 28

RESULT 17  
US-09-227-357-611  
Sequence 611, Application US/09227357  
Patent No. 6342581  
GENERAL INFORMATION:  
APPLICANT: Fischer et al.  
TITLE OF INVENTION: 123 Human Secreted Proteins  
FILE REFERENCE: P2010P1  
CURRENT APPLICATION NUMBER: US/09/227,357  
CURRENT FILING DATE: 1999-01-08  
EARLIER APPLICATION NUMBER: PCT/US98/13684  
EARLIER FILING DATE: 1998-07-07  
EARLIER APPLICATION NUMBER: 60/051,926  
EARLIER FILING DATE: 1997-07-08  
EARLIER APPLICATION NUMBER: 60/052,793  
EARLIER FILING DATE: 1997-07-08  
EARLIER APPLICATION NUMBER: 60/051,925  
EARLIER FILING DATE: 1997-07-08  
EARLIER APPLICATION NUMBER: 60/051,929  
EARLIER FILING DATE: 1997-07-08  
EARLIER APPLICATION NUMBER: 60/052,803

EARLIER FILING DATE: 1997-07-08  
EARLIER APPLICATION NUMBER: 60/052,732  
EARLIER FILING DATE: 1997-07-08  
EARLIER APPLICATION NUMBER: 60/051,931  
EARLIER FILING DATE: 1997-07-08  
EARLIER APPLICATION NUMBER: 60/051,932  
EARLIER FILING DATE: 1997-07-08  
EARLIER APPLICATION NUMBER: 60/051,916  
EARLIER FILING DATE: 1997-07-08  
EARLIER APPLICATION NUMBER: 60/051,930  
EARLIER FILING DATE: 1997-07-08  
EARLIER APPLICATION NUMBER: 60/051,918  
EARLIER FILING DATE: 1997-07-08  
EARLIER APPLICATION NUMBER: 60/051,920  
EARLIER FILING DATE: 1997-07-08  
EARLIER APPLICATION NUMBER: 60/052,733  
EARLIER FILING DATE: 1997-07-08  
EARLIER APPLICATION NUMBER: 60/052,795  
EARLIER FILING DATE: 1997-07-08  
EARLIER APPLICATION NUMBER: 60/051,919  
EARLIER FILING DATE: 1997-07-08  
EARLIER APPLICATION NUMBER: 60/051,928  
EARLIER FILING DATE: 1997-07-08  
EARLIER APPLICATION NUMBER: 60/055,722  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/055,723  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/055,948  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/055,949  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/055,953  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/055,950  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/055,947  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/055,964  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/056,360  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/055,684  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/055,984  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/055,954  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/058,785  
EARLIER FILING DATE: 1997-09-12  
EARLIER APPLICATION NUMBER: 60/058,664  
EARLIER FILING DATE: 1997-09-12  
EARLIER APPLICATION NUMBER: 60/058,660  
EARLIER FILING DATE: 1997-09-12  
EARLIER APPLICATION NUMBER: 60/058,661  
EARLIER FILING DATE: 1997-09-12  
NUMBER OF SEQ ID NOS: 672  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 611  
LENGTH: 63  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-227-357-611

Query Match 18.9%; Score 40; DB 4; Length 63;  
Best Local Similarity 47.1%; Pred. No. 50;  
Matches 8; Conservative 3; Mismatches 6; Indels 1; Gaps 0;

QY 5 ISENSLVAMDFSGQKSR 21  
Db 20 VSSNLSLFFPFGQKAR 36

RESULT 18

US-09-331-930A-2

; Sequence 2, Application US/09331930A  
; Patent No. 6436670  
; GENERAL INFORMATION:  
; APPLICANT: ZIMMET, PAUL Z.  
; TITLE OF INVENTION: A NOVEL GENE AND USES THEREFOR  
; FILE REFERENCE: 22975-20007.00  
; CURRENT APPLICATION NUMBER: US/09/331,930A  
; CURRENT FILING DATE: 1999-06-30  
; PRIOR APPLICATION NUMBER: PCT/AU98/00902  
; PRIOR FILING DATE: 1998-10-30  
; PRIOR APPLICATION NUMBER: AU PP0117/97  
; PRIOR FILING DATE: 1997-10-31  
; PRIOR APPLICATION NUMBER: AU PP0323/97  
; PRIOR FILING DATE: 1997-11-11  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 73  
; TYPE: PRT  
; ORGANISM: Unknown Organism  
; FEATURE:  
; OTHER INFORMATION: Description of Unknown Organism: Amino acid  
; OTHER INFORMATION: sequence for beacon from unknown organism  
US-09-331-930A-2

Query Match 18.6%; Score 39.5; DB 4; Length 73;  
Best Local Similarity 28.9%; Pred. No. 73;  
Matches 11; Conservative 6; Mismatches 16; Indels 5; Gaps 1;

QY 10 LVAMDFSGQKSRVIENTPEALS-----VAVEEGLAWRK 42  
DB 4 VVCNDRLGKVKRVKCNCTDDTIGDLKLLIAAQTGRWKK 41

RESULT 19

US-09-331-930A-19  
; Sequence 19, Application US/09331930A  
; Patent No. 6436670  
; GENERAL INFORMATION:  
; APPLICANT: ZIMMET, PAUL Z.  
; TITLE OF INVENTION: A NOVEL GENE AND USES THEREFOR  
; FILE REFERENCE: 22975-20007.00  
; CURRENT APPLICATION NUMBER: US/09/331,930A  
; CURRENT FILING DATE: 1999-06-30  
; PRIOR APPLICATION NUMBER: PCT/AU98/00902  
; PRIOR FILING DATE: 1998-10-30  
; PRIOR APPLICATION NUMBER: AU PP0117/97  
; PRIOR FILING DATE: 1997-10-31  
; PRIOR APPLICATION NUMBER: AU PP0323/97  
; PRIOR FILING DATE: 1997-11-11  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 19  
; LENGTH: 73  
; TYPE: PRT  
; ORGANISM: Unknown Organism  
; FEATURE:  
; OTHER INFORMATION: Description of Unknown Organism: Amino acid  
; OTHER INFORMATION: sequence for beacon from unknown organism  
US-09-331-930A-19

Query Match 18.6%; Score 39.5; DB 4; Length 73;  
Best Local Similarity 28.9%; Pred. No. 73;  
Matches 11; Conservative 6; Mismatches 16; Indels 5; Gaps 1;

QY 10 LVAMDFSGQKSRVIENTPEALS-----VAVEEGLAWRK 42  
DB 4 VVCNDRLGKVKRVKCNCTDDTIGDLKLLIAAQTGRWKK 41

RESULT 20

US-09-331-930A-20  
; Sequence 20, Application US/09331930A  
; Patent No. 6436670  
; GENERAL INFORMATION:  
; APPLICANT: ZIMMET, PAUL Z.  
; TITLE OF INVENTION: A NOVEL GENE AND USES THEREFOR  
; FILE REFERENCE: 22975-20007.00  
; CURRENT APPLICATION NUMBER: US/09/331,930A  
; CURRENT FILING DATE: 1999-06-30  
; PRIOR APPLICATION NUMBER: PCT/AU98/00902  
; PRIOR FILING DATE: 1998-10-30  
; PRIOR APPLICATION NUMBER: AU PP0117/97  
; PRIOR FILING DATE: 1997-10-31  
; PRIOR APPLICATION NUMBER: AU PP0323/97  
; PRIOR FILING DATE: 1997-11-11  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 20  
; LENGTH: 73  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-331-930A-20

Query Match 18.6%; Score 39.5; DB 4; Length 73;  
Best Local Similarity 28.9%; Pred. No. 73;  
Matches 11; Conservative 6; Mismatches 16; Indels 5; Gaps 1;

QY 10 LVAMDFSGQKSRVIENTPEALS-----VAVEEGLAWRK 42  
DB 4 VVCNDRLGKVKRVKCNCTDDTIGDLKLLIAAQTGRWKK 41

RESULT 21

US-09-331-930A-21  
; Sequence 21, Application US/09331930A  
; Patent No. 6436670  
; GENERAL INFORMATION:  
; APPLICANT: ZIMMET, PAUL Z.  
; TITLE OF INVENTION: A NOVEL GENE AND USES THEREFOR  
; FILE REFERENCE: 22975-20007.00  
; CURRENT APPLICATION NUMBER: US/09/331,930A  
; CURRENT FILING DATE: 1999-06-30  
; PRIOR APPLICATION NUMBER: PCT/AU98/00902  
; PRIOR FILING DATE: 1998-10-30  
; PRIOR APPLICATION NUMBER: AU PP0117/97  
; PRIOR FILING DATE: 1997-10-31  
; PRIOR APPLICATION NUMBER: AU PP0323/97  
; PRIOR FILING DATE: 1997-11-11  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 21  
; LENGTH: 73  
; TYPE: PRT  
; ORGANISM: Murine sp.  
US-09-331-930A-21

Query Match 18.6%; Score 39.5; DB 4; Length 73;  
Best Local Similarity 28.9%; Pred. No. 73;  
Matches 11; Conservative 6; Mismatches 16; Indels 5; Gaps 1;

QY 10 LVAMDFSGQKSRVIENTPEALS-----VAVEEGLAWRK 42  
DB 4 VVCNDRLGKVKRVKCNCTDDTIGDLKLLIAAQTGRWKK 41

RESULT 22

US-09-331-930A-24  
; Sequence 24, Application US/09331930A  
; Patent No. 6436670  
; GENERAL INFORMATION:

```

; APPLICANT: ZIMMET, PAUL Z.
; TITLE OF INVENTION: A NOVEL GENE AND USES THEREFOR
; FILE REFERENCE: 22975-2007.00
; CURRENT APPLICATION NUMBER: US/09/331.930A
; PRIOR FILING DATE: 1999-06-30
; PRIOR APPLICATION NUMBER: PCT/AU98/00902
; PRIOR FILING DATE: 1998-10-30
; PRIOR APPLICATION NUMBER: AU PP0117/97
; PRIOR FILING DATE: 1997-10-31
; PRIOR APPLICATION NUMBER: AU PP0323/97
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 24
; LENGTH: 73
; TYPE: PRT
; ORGANISM: Oryza sativa
US-09-331-930A-24

```

```

Query Match      18.6%; Score 39.5; DB 4; Length 73;
Best Local Similarity 28.9%; Pred. No. 73;
Matches 11; Conservative 6; Mismatches 16; Indels 5; Gaps 1;

```

```

QY 10 LVAMDFSGQSRVIENTPTALS-----VAVEEGLAWRK 42
DB 4 VVCDRLGKVRVKCNDTDTIGDLKKLIAAQTGTNRNK 41

```

```

RESULT 23
US-09-331-930A-25
; Sequence 25, Application US/09331930A
; Patent No. 6436670
; GENERAL INFORMATION:
; APPLICANT: ZIMMET, PAUL Z.

```

```

; TITLE OF INVENTION: A NOVEL GENE AND USES THEREFOR
; FILE REFERENCE: 22975-2007.00
; CURRENT APPLICATION NUMBER: US/09/331.930A
; PRIOR FILING DATE: 1999-06-30
; PRIOR APPLICATION NUMBER: PCT/AU98/00902
; PRIOR FILING DATE: 1998-10-30
; PRIOR APPLICATION NUMBER: AU PP0117/97
; PRIOR FILING DATE: 1997-10-31
; PRIOR APPLICATION NUMBER: AU PP0323/97
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 25
; LENGTH: 73
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-09-331-930A-25

```

```

Query Match      18.6%; Score 39.5; DB 4; Length 73;
Best Local Similarity 28.9%; Pred. No. 73;
Matches 11; Conservative 6; Mismatches 16; Indels 5; Gaps 1;

```

```

QY 10 LVAMDFSGQSRVIENTPTALS-----VAVEEGLAWRK 42
DB 4 VVCDRLGKVRVKCNDTDTIGDLKKLIAAQTGTNRNK 41

```

```

RESULT 24
US-09-100-804-30
; Sequence 30, Application US/09100804
; Patent No. 6068472
; GENERAL INFORMATION:
; APPLICANT: GONEZ, LEONEL JORGE
; APPLICANT: SARAS, JAN
; APPLICANT: CLAESSON-WELSH, LENA
; APPLICANT: HELDIN, CARL-HENRIK
; TITLE OF INVENTION: PRIMARY STRUCTURE AND FUNCTIONAL

```

```

; TITLE OF INVENTION: EXPRESSION OF NUCLEOTIDE SEQUENCES FOR NOVEL PROTEIN
; TITLE OF INVENTION: TYROSINE PHOSPHATASES
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.
; STREET: 600 ATLANTIC AVENUE
; CITY: BOSTON
; STATE: MASSACHUSETTS
; COUNTRY: USA
; ZIP: 02210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/100.804
; FILING DATE:
; CLASSIFICATION:

```

```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/596,291
; FILING DATE: 09-AUG-1996
; APPLICATION NUMBER: US 08/115,573
; FILING DATE: 01-SEP-1993

```

```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/09943
; FILING DATE: 01-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: GATES, EDWARD R.

```

```

; REGISTRATION NUMBER: 31,616
; REFERENCE/DOCKET NUMBER: L0461/7003
; TELEPHONE: 617-720-3500
; TELEFAX: 617-720-2441
; TELEX:

```

```

; INFORMATION FOR SEQ ID NO: 30:

```

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 68 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-09-100-804-30

```

```

Query Match      18.4%; Score 39; DB 3; Length 68;
Best Local Similarity 30.4%; Pred. No. 79;
Matches 14; Conservative 9; Mismatches 19; Indels 4; Gaps 1;

```

```

QY 2 MRSISENSLVAMDFSGQSRVI----ENPTALSVAVEEGLAWRK 43
DB 16 VKRISQDSLAARDGDIQGDVVKINGTVTNNSLTDKTLIERK 61

```

```

RESULT 25

```

```

US-09-081-320-9
; Sequence 9, Application US/09081320
; Patent No. 6093544
; GENERAL INFORMATION:

```

```

; APPLICANT: Gonsalves, Dennis
; APPLICANT: Meng, Baozhong

```

```

; TITLE OF INVENTION: RUPESTRIS STEM PITTING ASSOCIATED VIRUS
; TITLE OF INVENTION: NUCLEIC ACIDS, PROTEINS, AND THEIR USES
; NUMBER OF SEQUENCES: 54

```

```

; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
; STREET: Clinton Square, P.O. Box 1051
; CITY: Rochester
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 14603
; COMPUTER READABLE FORM:

```

Fri Mar 28 12:13:32 2003

us-09-936-697-5.max.rai

Page 16

;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.30  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/09/081,320  
;; FILING DATE:  
;; CLASSIFICATION:  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 60/047,147  
;; FILING DATE: 20-MAY-1997  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 60/069,902  
;; FILING DATE: 17-DEC-1997  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Goldman, Michael L.  
;; REGISTRATION NUMBER: 30,727  
;; REFERENCE/DOCKET NUMBER: 19603/1722  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (716) 263-1304  
;; TELEFAX: (716) 263-1600  
;; INFORMATION FOR SEQ ID NO: 9:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 80 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS:  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
US-09-081-320-9

Query Match 18.4%; Score 39; DB 3; Length 80;  
Best Local Similarity 46.2%; Pred No 98;  
Matches 12; Conservative 5; Mismatches 7; Indels 2; Gaps 1;  
QY 19 KSRVIEN--PTEALSVAVEGLAWRK 42  
Db 40 ESILINCSPSEALASTVKEVLGLK 65

Search completed: March 28, 2003, 09:09:31  
Job time : 30.5276 secs

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 28, 2003, 08:55:25 ; Search time 33.8583 Seconds  
(without alignments)  
169.228 Million cell updates/sec

Sequence: 1 PMSISENSLVAMDFSGK.....ENPTALSVAVEGLAWRK 43

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues  
Total number of hits satisfying chosen parameters: 160864

Minimum DB seq length: 40  
Maximum DB seq length: 90

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 1000 summaries

Result No.	Score	Query Match	Length	ID	Description
1	212	100.0	43	21	Peptide derived fr
2	212	100.0	84	21	Peptide derived fr
3	205	96.7	43	21	Peptide derived fr
4	205	96.7	84	21	Peptide derived fr
5	159	79.7	43	21	Peptide derived fr
6	159	79.7	84	21	Peptide derived fr
7	162	76.4	43	21	Peptide derived fr
8	162	76.4	84	21	Peptide derived fr
9	161	75.9	43	21	Peptide derived fr
10	161	75.9	84	21	Peptide derived fr

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	212	100.0	43	21	Peptide derived fr
2	212	100.0	84	21	Peptide derived fr
3	205	96.7	43	21	Peptide derived fr
4	205	96.7	84	21	Peptide derived fr
5	159	79.7	43	21	Peptide derived fr
6	159	79.7	84	21	Peptide derived fr
7	162	76.4	43	21	Peptide derived fr
8	162	76.4	84	21	Peptide derived fr
9	161	75.9	43	21	Peptide derived fr
10	161	75.9	84	21	Peptide derived fr

11	159	75.0	43	21	Peptide derived fr
12	159	75.0	43	21	Peptide derived fr
13	159	75.0	80	21	Peptide derived fr
14	159	75.0	80	21	Peptide derived fr
15	49	23.1	67	23	Human ORFX protein
16	48	22.6	72	22	Novel human secret
17	47.5	22.4	73	22	Human colon cancer
18	45.5	21.5	73	22	Human ORFX protein
19	45	21.2	84	23	Human ORFX protein
20	44.5	21.0	72	21	Human secreted pro
21	44.5	21.0	74	22	Novel central nerv
22	44.5	21.0	74	22	Novel central nerv
23	44.5	21.0	79	22	Protonibacterium
24	44.5	21.0	90	22	Protonibacterium
25	44	20.8	47	22	Human reproductive
26	44	20.8	62	22	Novel human diagno
27	44	20.8	80	20	RSPV strain RSP47
28	44	20.8	82	22	Drosophila melanog
29	42	19.8	45	22	Human immune/haema
30	42	19.8	63	22	Human peptide #979
31	42	19.8	63	22	Peptide #1011 enco
32	42	19.8	63	22	Protein #963 enco
33	42	19.8	63	22	Human brain expres
34	42	19.8	63	22	Human bone marrow
35	42	19.8	63	22	Peptide #981 enco
36	42	19.8	63	22	Peptide #1002 enco
37	42	19.8	63	22	Peptide #955 enco
38	42	19.8	63	22	Human immune/haema
39	42	19.8	64	22	Human immune/haema
40	42	19.8	72	22	C glutamic prote
41	42	19.8	81	20	Chlamydia pneumoni
42	42	19.8	87	23	Human ORFX protein
43	42	19.8	89	22	Human polypeptide
44	41.5	19.6	51	22	Human reproductive
45	41.5	19.6	71	22	Protonibacterium
46	41.5	19.6	82	22	Human nervous syst
47	41	19.3	53	22	Protonibacterium
48	41	19.3	63	22	Human ORFX protein
49	41	19.3	89	22	Protonibacterium
50	40.5	19.1	51	21	Bacteriophage Dp-1
51	40.5	19.1	56	21	SH3 domain from P8
52	40.5	19.1	66	21	Arabidopsis thalia
53	40.5	19.1	74	21	Human secreted pro
54	40.5	19.1	82	21	Human secreted pro
55	40	18.9	49	22	Peptide #10267 enc
56	40	18.9	49	22	Protein #8053 enco
57	40	18.9	49	22	Human brain expres
58	40	18.9	49	22	Human bone marrow
59	40	18.9	49	22	Peptide #10610 enc
60	40	18.9	49	22	Human peptide enco
61	40	18.9	57	21	Human secreted pro
62	40	18.9	63	20	Fragment of human
63	40	18.9	70	19	H. pylori GHPO 138
64	40	18.9	71	22	Protonibacterium
65	40	18.9	71	22	Novel human colon
66	40	18.9	71	22	Human digestive sy
67	40	18.9	72	21	Arabidopsis thalia
68	40	18.9	74	21	Arabidopsis thalia
69	39.5	18.6	54	20	Magnaporthe grisea
70	39.5	18.6	61	23	Lactococcus lactis
71	39.5	18.6	68	22	Human bone marrow
72	39.5	18.6	68	22	Peptide #7275 enco
73	39.5	18.6	68	22	Human peptide enco
74	39.5	18.6	73	20	P. obesus beaon p
75	39.5	18.6	73	21	Israeli sand rat b
76	39.5	18.6	73	22	Human polyptide
77	39.5	18.6	78	22	Human digestive sy
78	39.5	18.6	78	22	Human liver associ
79	39.5	18.6	78	22	Human liver antige
80	39.5	18.6	83	22	Novel human diagno
81	39.5	18.6	86	21	Arabidopsis thalia
82	39.5	18.6	89	22	Enterococcus faeca
83	39	18.4	40	22	ERA binding domain















960 33 15.6 84 21 AAG13566 Arabidopsis thalia  
961 33 15.6 84 23 AAU97686 RNA3' terminal pho  
962 33 15.6 85 22 AAO11826 Human polypeptide  
963 33 15.6 86 22 ABB30838 Peptide #3489 enco  
964 33 15.6 86 22 ABB36016 Peptide #3522 enco  
965 33 15.6 86 22 ABB21431 Protein #3420 enco  
966 33 15.6 86 22 AAM56809 Human brain expres  
967 33 15.6 86 22 AAM56909 Human bone marrow  
968 33 15.6 86 22 AAM17022 Peptide #3456 enco  
969 33 15.6 86 22 AAM29313 Peptide #3550 enco  
970 33 15.6 86 22 AAM04730 Peptide #3412 enco  
971 33 15.6 86 22 AAG38793 Human peptide enco  
972 33 15.6 87 21 AAG35303 Zea mays protein f  
973 33 15.6 87 22 AAU67508 Propionibacterium  
974 33 15.6 88 19 AAW50153 Taranula venom pr  
975 33 15.6 88 21 AAB11531 SEN virus protein  
976 33 15.6 88 22 AAU67025 Propionibacterium  
977 33 15.6 88 22 AAM88974 Human immune/haema  
978 33 15.6 90 21 AAG14932 Arabidopsis thalia  
979 33 15.6 90 21 AAG15442 Arabidopsis thalia  
980 33 15.6 90 21 AAG26831 Zea mays protein f  
981 33 15.6 90 21 AAG35425 Arabidopsis thalia  
982 33 15.6 90 21 AAG49292 Arabidopsis thalia  
983 33 15.6 90 22 AAU53146 Propionibacterium  
984 33 15.6 90 22 AAU18439 Human endocrine po  
985 32.5 15.3 42 22 AAU22114 Human cardiovascular  
986 32.5 15.3 43 22 AAO11974 Human polypeptide  
987 32.5 15.3 45 13 AAR29176 Astrovirus serotyp  
988 32.5 15.3 46 22 AAO10620 Human polypeptide  
989 32.5 15.3 47 13 AAB35397 Lactococcus lactis  
990 32.5 15.3 47 13 AAR20763 Peptide 13 based o  
991 32.5 15.3 49 22 ABB44269 Peptide #11775 enc  
992 32.5 15.3 49 22 AAM65305 Protein #9135 enco  
993 32.5 15.3 49 22 AAM78004 Human bone marrow  
994 32.5 15.3 49 22 AAM78004 Human bone marrow  
995 32.5 15.3 49 22 AAM78004 Peptide #9328 enco  
996 32.5 15.3 49 22 AAM78004 Peptide #12257 enc  
997 32.5 15.3 49 23 AAG47017 Human peptide enco  
998 32.5 15.3 50 5 AAP40503 Sequence of the CN  
999 32.5 15.3 50 21 AAG55355 Arabidopsis thalia  
1000 32.5 15.3 50 21 AAG55743 Arabidopsis thalia

# ALIGNMENTS

RESULT 1  
ID AAB18941 standard; peptide; 43 AA.  
AC AAB18941;  
DT 08-FEB-2001 (first entry)  
DE Peptide derived from the PIR or PIR-SH2 domain of Grb7 protein.  
KW Phosphorylated insulin receptor interacting region; Grb7 family protein;  
KW insulin receptor; tyrosine kinase; insulin; insulin-associated disease;  
KW diabetes; obesity; polycystic ovarian syndrome; syndrome X.  
OS Homo sapiens.  
PN WO200055634-A1  
PS 21-SEP-2000.  
PF 14-MAR-2000; 2000WO-FR00613.  
PR 15-MAR-1999; 99FR-0003159.  
PA (CNRS ) CNRS CENT NAT RECH SCI.  
PI Burnol A, Perdereau D, Kasus-Jacobi A, Berezat V, Girard J;

XX WPI; 2000-587566/55.  
XX Fragments of Grb family proteins to identify compounds are useful in  
XX treating insulin-associated diseases, particularly diabetes and obesity  
PT Claim 2; Page 25; 46pp; French.  
PS B18937-64 represent the PIR (phosphorylated insulin receptor interacting  
CC region) or PIR-SH2 (Src homology 2) domains of a Grb7 family protein.  
CC PIR is the actual binding region but its effect is about 10 times  
CC greater in presence of SH2 (which by itself is inactive). Agents that  
CC affect binding between the peptides and the insulin receptor can  
CC stimulate or inhibit tyrosine kinase activity of the receptor. The  
CC peptides are used for screening molecules for ability to treat diseases  
CC in which insulin is implicated. The peptides are used to identify agents  
CC that are potentially useful for treating insulin-associated diseases,  
CC particularly diabetes and obesity but also polycystic ovarian syndrome  
CC and syndrome X.  
XX Sequence 43 AA;  
SQ Query Match 100.0%; Score 212; DB 21; Length 43;  
Best Local Similarity 100.0%; Pred. No. 7.9e-25;  
Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 PMRSISENSLVAMDFSGQSRVIENTEALSVAVEEGLAWRKK 43  
Db 1 PMRSISENSLVAMDFSGQSRVIENTEALSVAVEEGLAWRKK 43  
RESULT 2  
AAB18942  
ID AAB18942 standard; peptide; 84 AA.  
AC AAB18942;  
DT 08-FEB-2001 (first entry)  
DE Peptide derived from the PIR or PIR-SH2 domain of Grb7 protein.  
KW Phosphorylated insulin receptor interacting region; Grb7 family protein;  
KW insulin receptor; tyrosine kinase; insulin; insulin-associated disease;  
KW diabetes; obesity; polycystic ovarian syndrome; syndrome X.  
OS Homo sapiens.  
PN WO200055634-A1.  
PS 21-SEP-2000.  
PF 14-MAR-2000; 2000WO-FR00613.  
PR 15-MAR-1999; 99FR-0003159.  
PA (CNRS ) CNRS CENT NAT RECH SCI.  
PI Burnol A, Perdereau D, Kasus-Jacobi A, Berezat V, Girard J;  
XX WPI; 2000-587566/55.  
XX Fragments of Grb family proteins to identify compounds are useful in  
XX treating insulin-associated diseases, particularly diabetes and obesity  
PT Claim 2; Page 26; 46pp; French.  
PS B18937-64 represent the PIR (phosphorylated insulin receptor interacting  
CC region) or PIR-SH2 (Src homology 2) domains of a Grb7 family protein.  
CC PIR is the actual binding region but its effect is about 10 times  
CC greater in presence of SH2 (which by itself is inactive). Agents that  
CC affect binding between the peptides and the insulin receptor can

CC stimulate or inhibit tyrosine kinase activity of the receptor. The  
 CC peptides are used for screening molecules for ability to treat diseases  
 CC in which insulin is implicated. The peptides are used to identify agents  
 CC that are potentially useful for treating insulin-associated diseases,  
 CC particularly diabetes and obesity but also polycystic ovarian syndrome  
 CC and syndrome X.  
 XX Sequence 84 AA;  
 XX DB 21; Length 84;  
 Query Match 100.0%; Score 212; DB 21; Length 84;  
 Best Local Similarity 100.0%; Pred. No. 2e-24;  
 Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 PMRSSENSLVAMDFSGQKSRVIENTEALSVAVEGLAWRKK 43  
 DB 13 PMRSSENSLVAMDFSGQKSRVIENTEALSVAVEGLAWRKK 55  
 RESULT 3  
 AAB18937  
 ID AAB18937 standard; peptide; 43 AA.  
 AC AAB18937;  
 DT 08-FEB-2001 (first entry)  
 DE Peptide derived from the PIR or PIR-SH2 domain of Grb7 protein.  
 KW phosphorylated insulin receptor interacting region; Grb7 family protein;  
 KW insulin receptor; tyrosine kinase; insulin; insulin-associated disease;  
 KW diabetes; obesity; polycystic ovarian syndrome; syndrome X.  
 XX Rattus sp.  
 OS WO200055634-A1.  
 PN 21-SEP-2000.  
 PD 14-MAR-2000; 2000WO-FR00613.  
 PF 15-MAR-1999; 99FR-0003159.  
 PR (CNRS ) CNRS CENT NAT RECH SCI.  
 PA Burnol A, Perdureau D, Kasus-Jacobi A, Bereziat V, Girard J;  
 PI WPI; 2000-587566/55.  
 PS Fragments of Grb family proteins to identify compounds are useful in  
 PT treating insulin-associated diseases, particularly diabetes and obesity  
 PT Claim 2; Page 23; 46pp; French.  
 XX Bl8937-64 represent the PIR (phosphorylated insulin receptor interacting  
 CC region) or PIR-SH2 (Src homology 2) domains of a Grb7 family protein.  
 CC PIR is the actual binding region but its effect is about 10 times  
 CC greater in presence of SH2 (which by itself is inactive). Agents that  
 CC affect binding between the peptides and the insulin receptor can  
 CC stimulate or inhibit tyrosine kinase activity of the receptor. The  
 CC peptides are used for screening molecules for ability to treat diseases  
 CC in which insulin is implicated. The peptides are used to identify agents  
 CC that are potentially useful for treating insulin-associated diseases,  
 CC particularly diabetes and obesity but also polycystic ovarian syndrome  
 CC and syndrome X.  
 XX Sequence 43 AA;  
 XX DB 21; Length 43;  
 Query Match 96.7%; Score 205; DB 21; Length 43;  
 Best Local Similarity 93.0%; Pred. No. 9.2e-24;  
 Matches 40; Conservative 3; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 PMRSSENSLVAMDFSGQKSRVIENTEALSVAVEGLAWRKK 43

DB 1 PMRSSENSLVAMDFSGQKTRVIDNPTSEALSVAVEGLAWRKK 43  
 RESULT 4  
 AAB18938  
 ID AAB18938 standard; peptide; 84 AA.  
 AC AAB18938;  
 DT 08-FEB-2001 (first entry)  
 DE Peptide derived from the PIR or PIR-SH2 domain of Grb7 protein.  
 KW phosphorylated insulin receptor interacting region; Grb7 family protein;  
 KW insulin receptor; tyrosine kinase; insulin; insulin-associated disease;  
 KW diabetes; obesity; polycystic ovarian syndrome; syndrome X.  
 XX Rattus sp.  
 OS WO200055634-A1.  
 PN 21-SEP-2000.  
 PD 14-MAR-2000; 2000WO-FR00613.  
 PF 15-MAR-1999; 99FR-0003159.  
 PR (CNRS ) CNRS CENT NAT RECH SCI.  
 PA Burnol A, Perdureau D, Kasus-Jacobi A, Bereziat V, Girard J;  
 PI WPI; 2000-587566/55.  
 PS Fragments of Grb family proteins to identify compounds are useful in  
 PT treating insulin-associated diseases, particularly diabetes and obesity  
 PT Claim 2; Page 23-24; 46pp; French.  
 XX Bl8937-64 represent the PIR (phosphorylated insulin receptor interacting  
 CC region) or PIR-SH2 (Src homology 2) domains of a Grb7 family protein.  
 CC PIR is the actual binding region but its effect is about 10 times  
 CC greater in presence of SH2 (which by itself is inactive). Agents that  
 CC affect binding between the peptides and the insulin receptor can  
 CC stimulate or inhibit tyrosine kinase activity of the receptor. The  
 CC peptides are used for screening molecules for ability to treat diseases  
 CC in which insulin is implicated. The peptides are used to identify agents  
 CC that are potentially useful for treating insulin-associated diseases,  
 CC particularly diabetes and obesity but also polycystic ovarian syndrome  
 CC and syndrome X.  
 XX Sequence 84 AA;  
 XX DB 21; Length 84;  
 Query Match 96.7%; Score 205; DB 21; Length 84;  
 Best Local Similarity 93.0%; Pred. No. 2.3e-23;  
 Matches 40; Conservative 3; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 PMRSSENSLVAMDFSGQKSRVIENTEALSVAVEGLAWRKK 43  
 DB 13 PMRSSENSLVAMDFSGQKTRVIDNPTSEALSVAVEGLAWRKK 55  
 RESULT 5  
 AAB18949  
 ID AAB18949 standard; peptide; 43 AA.  
 AC AAB18949;  
 DT 08-FEB-2001 (first entry)  
 DE Peptide derived from the PIR or PIR-SH2 domain of Grb7 protein.  
 XX

KW Phosphorylated insulin receptor interacting region; Grb7 family protein;  
 KW insulin receptor; tyrosine kinase; insulin; insulin-associated disease;  
 XX diabetes; obesity; polycystic ovarian syndrome; syndrome X.  
 OS Homo sapiens.

XX WO2000055634-A1.  
 XX 21-SEP-2000.  
 XX 14-MAR-2000; 2000WO-FR00613.  
 XX 15-MAR-1999; 99FR-0003159.

XX (CNRS ) CNRS CENT NAT RECH SCI.  
 XX Burnol A, Perdureau D, Kasus-Jacobi A, Bereziat V, Girard J;  
 XX WPI; 2000-587566/55.

XX Fragments of Grb family proteins to identify compounds are useful in  
 XX treating insulin-associated diseases, particularly diabetes and obesity  
 XX -

XX Claim 2; Page 30; 46pp; French.

XX BI8937-64 represent the PIR (phosphorylated insulin receptor interacting  
 XX region) or PIR-SH2 (Src homology 2) domains of a Grb7 family protein.  
 XX PIR is the actual binding region but its effect is about 10 times  
 XX greater in presence of SH2 (which by itself is inactive). Agents that  
 XX affect binding between the peptides and the insulin receptor can  
 XX stimulate or inhibit tyrosine kinase activity of the receptor. The  
 XX peptides are used for screening molecules for ability to treat diseases  
 XX in which insulin is implicated. The peptides are used to identify agents  
 XX that are potentially useful for treating insulin-associated diseases,  
 XX particularly diabetes and obesity but also polycystic ovarian syndrome  
 XX and syndrome X.

XX Query Match 79.7%; Score 169; DB 21; Length 43;  
 XX Best Local Similarity 76.7%; Pred. No. 2.9e-18;  
 XX Matches 33; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 1 PMRSISENSLVAMDFSGOKSRVIENTEALSVAVEGLAWRKK 43  
 DB 1 PVRVSENSLVAMDFSGOTGRVIENTEALSVAVEGLAWRKK 43

RESULT 6  
 AAB18950  
 ID AAB18950 standard; peptide: 82 AA.

AC AAB18950;  
 XX 08-FEB-2001 (first entry)

DE Peptide derived from the PIR or PIR-SH2 domain of Grb7 protein.

XX Phosphorylated insulin receptor interacting region; Grb7 family protein;  
 KW insulin receptor; tyrosine kinase; insulin; insulin-associated disease;  
 KW diabetes; obesity; polycystic ovarian syndrome; syndrome X.

XX Homo sapiens.  
 XX WO2000055634-A1.  
 XX 21-SEP-2000.  
 XX 14-MAR-2000; 2000WO-FR00613.  
 XX 15-MAR-1999; 99FR-0003159.

XX (CNRS ) CNRS CENT NAT RECH SCI.  
 XX Burnol A, Perdureau D, Kasus-Jacobi A, Bereziat V, Girard J;  
 XX WPI; 2000-587566/55.

XX Fragments of Grb family proteins to identify compounds are useful in  
 XX treating insulin-associated diseases, particularly diabetes and obesity  
 XX -

XX Claim 2; Page 30; 46pp; French.

XX BI8937-64 represent the PIR (phosphorylated insulin receptor interacting  
 XX region) or PIR-SH2 (Src homology 2) domains of a Grb7 family protein.  
 XX PIR is the actual binding region but its effect is about 10 times  
 XX greater in presence of SH2 (which by itself is inactive). Agents that  
 XX affect binding between the peptides and the insulin receptor can  
 XX stimulate or inhibit tyrosine kinase activity of the receptor. The  
 XX peptides are used for screening molecules for ability to treat diseases  
 XX in which insulin is implicated. The peptides are used to identify agents  
 XX that are potentially useful for treating insulin-associated diseases,  
 XX particularly diabetes and obesity but also polycystic ovarian syndrome  
 XX and syndrome X.

XX Query Match 79.7%; Score 169; DB 21; Length 43;  
 XX Best Local Similarity 76.7%; Pred. No. 2.9e-18;  
 XX Matches 33; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 1 PMRSISENSLVAMDFSGOKSRVIENTEALSVAVEGLAWRKK 43  
 DB 1 PVRVSENSLVAMDFSGOTGRVIENTEALSVAVEGLAWRKK 43

RESULT 6  
 AAB18950  
 ID AAB18950 standard; peptide: 82 AA.

AC AAB18950;  
 XX 08-FEB-2001 (first entry)

DE Peptide derived from the PIR or PIR-SH2 domain of Grb7 protein.

XX Phosphorylated insulin receptor interacting region; Grb7 family protein;  
 KW insulin receptor; tyrosine kinase; insulin; insulin-associated disease;  
 KW diabetes; obesity; polycystic ovarian syndrome; syndrome X.

XX Homo sapiens.  
 XX WO2000055634-A1.  
 XX 21-SEP-2000.  
 XX 14-MAR-2000; 2000WO-FR00613.  
 XX 15-MAR-1999; 99FR-0003159.

XX (CNRS ) CNRS CENT NAT RECH SCI.  
 XX Burnol A, Perdureau D, Kasus-Jacobi A, Bereziat V, Girard J;  
 XX WPI; 2000-587566/55.

XX Fragments of Grb family proteins to identify compounds are useful in  
 XX treating insulin-associated diseases, particularly diabetes and obesity  
 XX -

XX Claim 2; Page 34; 46pp; French.

XX BI8937-64 represent the PIR (phosphorylated insulin receptor interacting  
 XX region) or PIR-SH2 (Src homology 2) domains of a Grb7 family protein.

PA (CNRS ) CNRS CENT NAT RECH SCI.  
 XX Burnol A, Perdureau D, Kasus-Jacobi A, Bereziat V, Girard J;  
 XX WPI; 2000-587566/55.

XX Fragments of Grb family proteins to identify compounds are useful in  
 XX treating insulin-associated diseases, particularly diabetes and obesity  
 XX -

XX Claim 2; Page 30; 46pp; French.

XX BI8937-64 represent the PIR (phosphorylated insulin receptor interacting  
 XX region) or PIR-SH2 (Src homology 2) domains of a Grb7 family protein.  
 XX PIR is the actual binding region but its effect is about 10 times  
 XX greater in presence of SH2 (which by itself is inactive). Agents that  
 XX affect binding between the peptides and the insulin receptor can  
 XX stimulate or inhibit tyrosine kinase activity of the receptor. The  
 XX peptides are used for screening molecules for ability to treat diseases  
 XX in which insulin is implicated. The peptides are used to identify agents  
 XX that are potentially useful for treating insulin-associated diseases,  
 XX particularly diabetes and obesity but also polycystic ovarian syndrome  
 XX and syndrome X.

XX Query Match 79.7%; Score 169; DB 21; Length 82;  
 XX Best Local Similarity 76.7%; Pred. No. 7e-18;  
 XX Matches 33; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 1 PMRSISENSLVAMDFSGOKSRVIENTEALSVAVEGLAWRKK 43  
 DB 13 PVRVSENSLVAMDFSGOTGRVIENTEALSVAVEGLAWRKK 55

RESULT 7  
 AAB18957  
 ID AAB18957 standard; peptide: 43 AA.

AC AAB18957;  
 XX 08-FEB-2001 (first entry)

DE Peptide derived from the PIR or PIR-SH2 domain of Grb7 protein.

XX Phosphorylated insulin receptor interacting region; Grb7 family protein;  
 KW insulin receptor; tyrosine kinase; insulin; insulin-associated disease;  
 KW diabetes; obesity; polycystic ovarian syndrome; syndrome X.

XX Homo sapiens.  
 XX WO2000055634-A1.  
 XX 21-SEP-2000.  
 XX 14-MAR-2000; 2000WO-FR00613.  
 XX 15-MAR-1999; 99FR-0003159.

XX (CNRS ) CNRS CENT NAT RECH SCI.  
 XX Burnol A, Perdureau D, Kasus-Jacobi A, Bereziat V, Girard J;  
 XX WPI; 2000-587566/55.

XX Fragments of Grb family proteins to identify compounds are useful in  
 XX treating insulin-associated diseases, particularly diabetes and obesity  
 XX -

XX Claim 2; Page 34; 46pp; French.

XX BI8937-64 represent the PIR (phosphorylated insulin receptor interacting  
 XX region) or PIR-SH2 (Src homology 2) domains of a Grb7 family protein.

CC PIR is the actual binding region but its effect is about 10 times  
 CC greater in presence of SH2 (which by itself is inactive). Agents that  
 CC affect binding between the peptides and the insulin receptor can.  
 CC stimulate or inhibit tyrosine kinase activity of the receptor. The  
 CC peptides are used for screening molecules for ability to treat diseases  
 CC in which insulin is implicated. The peptides are used to identify agents  
 CC that are potentially useful for treating insulin-associated diseases,  
 CC particularly diabetes and obesity but also polycystic ovarian syndrome  
 CC and syndrome X.

XX SQ Sequence 43 AA;

Query Match 76.4%; Score 162; DB 21; Length 43;

Best Local Similarity 74.4%; Pred. No. 3.3e-17;

Matches 32; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

Oy 1 PMRSISENSLVAMDFSGKSRVIENTPEALSVAVEGLAWRK 43

Db 1 PLRSASDNTLVAMDFSGHAGRVIENTPREALSVALEEAQAWRK 43

RESULT 8

AAB18958

ID AAB18958 standard; peptide: 80 AA.

XX AC

XX AAB18958;

XX DT

08-FEB-2001 (first entry)

XX DE

Peptide derived from the PIR or PIR-SH2 domain of Grb7 protein.

XX KW

Phosphorylated insulin receptor interacting region; Grb7 family protein;

insulin receptor; tyrosine kinase; insulin; insulin-associated disease;

XX KW

diabetes; obesity; polycystic ovarian syndrome; syndrome X.

XX OS

Homo sapiens.

XX FN

WO200055634-A1.

XX PD

21-SEP-2000.

XX PF

14-MAR-2000; 2000WO-FR00613.

XX PR

15-MAR-1999; 99FR-0003159.

XX PA

(CNRS ) CNRS CENT NAT RECH SCI.

XX PI

Burnol A, Perdureau D, Kasus-Jacobi A, Bereziat V, Girard J;

XX DR

WPI; 2000-587566/55.

XX PT

Fragments of Grb family proteins to identify compounds are useful in

XX PT

treating insulin-associated diseases, particularly diabetes and obesity

XX PS

Claim 2; Page 34-35; 46pp; French.

XX CC

BI8937-64 represent the PIR (phosphorylated insulin receptor interacting

region) or PIR-SH2 (src homology 2) domains of a Grb7 family protein.

CC PIR is the actual binding region but its effect is about 10 times

greater in presence of SH2 (which by itself is inactive). Agents that

affect binding between the peptides and the insulin receptor can

stimulate or inhibit tyrosine kinase activity of the receptor. The

peptides are used for screening molecules for ability to treat diseases

in which insulin is implicated. The peptides are used to identify agents

that are potentially useful for treating insulin-associated diseases,

particularly diabetes and obesity but also polycystic ovarian syndrome

CC and syndrome X.

XX SQ

Sequence 80 AA;

Query Match 76.4%; Score 162; DB 21; Length 80;

Best Local Similarity 74.4%; Pred. No. 7.9e-17;

Matches 32; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

Oy 1 PMRSISENSLVAMDFSGKSRVIENTPEALSVAVEGLAWRK 43

Db 13 PLRSASDNTLVAMDFSGHAGRVIENTPREALSVALEEAQAWRK 55

RESULT 9

AAB18945

ID AAB18945 standard; peptide: 43 AA.

XX AC

XX AAB18945;

XX DT

08-FEB-2001 (first entry)

XX DE

Peptide derived from the PIR or PIR-SH2 domain of Grb7 protein.

XX KW

Phosphorylated insulin receptor interacting region; Grb7 family protein;

insulin receptor; tyrosine kinase; insulin; insulin-associated disease;

XX KW

diabetes; obesity; polycystic ovarian syndrome; syndrome X.

XX OS

Mus musis.

XX FN

WO200055634-A1.

XX PD

21-SEP-2000.

XX PF

14-MAR-2000; 2000WO-FR00613.

XX PR

15-MAR-1999; 99FR-0003159.

XX PA

(CNRS ) CNRS CENT NAT RECH SCI.

XX PI

Burnol A, Perdureau D, Kasus-Jacobi A, Bereziat V, Girard J;

XX DR

WPI; 2000-587566/55.

XX PT

Fragments of Grb family proteins to identify compounds are useful in

XX PT

treating insulin-associated diseases, particularly diabetes and obesity

XX PS

Claim 2; Page 27-28; 46pp; French.

XX CC

BI8937-64 represent the PIR (phosphorylated insulin receptor interacting

region) or PIR-SH2 (src homology 2) domains of a Grb7 family protein.

CC PIR is the actual binding region but its effect is about 10 times

greater in presence of SH2 (which by itself is inactive). Agents that

affect binding between the peptides and the insulin receptor can

stimulate or inhibit tyrosine kinase activity of the receptor. The

peptides are used for screening molecules for ability to treat diseases

in which insulin is implicated. The peptides are used to identify agents

that are potentially useful for treating insulin-associated diseases,

particularly diabetes and obesity but also polycystic ovarian syndrome

CC and syndrome X.

XX SQ

Sequence 43 AA;

Query Match 75.9%; Score 161; DB 21; Length 43;

Best Local Similarity 78.0%; Pred. No. 4.8e-17;

Matches 32; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

Oy 1 PMRSISENSLVAMDFSGKSRVIENTPEALSVAVEGLAWRK 41

Db 1 PMRSISENSLVAMDFSGQIGRVINDNPAQSAALEEGHAWRK 41

RESULT 10

AAB18946

ID AAB18946 standard; peptide: 82 AA.

XX AC

XX AAB18946;

XX DT

08-FEB-2001 (first entry)





XX BI8937-64 represent the PIR (phosphorylated insulin receptor interacting  
CC region) or PIR-SH2 (Src homology 2) domains of a Grb7 family protein.  
CC PIR is the actual binding region but its effect is about 10 times  
CC greater in presence of SH2 (which by itself is inactive). Agents that  
CC affect binding between the peptides and the insulin receptor can  
CC stimulate or inhibit tyrosine kinase activity of the receptor. The  
CC peptides are used for screening molecules for ability to treat diseases  
CC in which insulin is implicated. The peptides are used to identify agents  
CC that are potentially useful for treating insulin-associated diseases,  
CC particularly diabetes and obesity but also polycystic ovarian syndrome  
CC and syndrome X.

XX Sequence 43 AA:  
XX  
XX Query Match 75.0%; Score 159; DB 21; Length 43;  
XX Best Local Similarity 69.8%; Pred. No. 9.6e-17;  
XX Matches 30; Conservative 6; Mismatches 7; Indels 0; Gaps 0;

OY 1 PMSISENSLVAMDFSGOKSRVIENTPEALSVAVEGLAWRK 43  
DB 1 PLSVSDNTLVAMDFSGHAGRVINDNPREALSAAEQAQRKK 43

RESULT 13  
AAB18954  
ID AAB18954 standard; peptide: 80 AA.

XX AAB18954;  
XX  
XX 08-FEB-2001 (first entry)  
XX  
XX Peptide derived from the PIR or PIR-SH2 domain of Grb7 protein.

XX Phosphorylated insulin receptor interacting region; Grb7 family protein;  
KW insulin receptor; tyrosine kinase; insulin; insulin-associated disease;  
KW diabetes; obesity; polycystic ovarian syndrome; syndrome X.

XX Rattus sp.

XX WO200055634-A1.

XX 21-SEP-2000.

XX 14-MAR-2000; 2000WO-FR00613.

XX 15-MAR-1999; 99FR-0003159.

XX (CNRS ) CNRS CENT NAT RECH SCI.

PI Burnol A, Perdereau D, Kasus-Jacobi A, Bereziat V, Girard J;

DR WPI; 2000-587566/55.

XX Fragments of Grb family proteins to identify compounds are useful in  
PT treating insulin-associated diseases, particularly diabetes and obesity  
PT

PS Claim 2; Page 32; 46pp; French.

XX BI8937-64 represent the PIR (phosphorylated insulin receptor interacting  
CC region) or PIR-SH2 (Src homology 2) domains of a Grb7 family protein.  
CC PIR is the actual binding region but its effect is about 10 times  
CC greater in presence of SH2 (which by itself is inactive). Agents that  
CC affect binding between the peptides and the insulin receptor can  
CC stimulate or inhibit tyrosine kinase activity of the receptor. The  
CC peptides are used for screening molecules for ability to treat diseases  
CC in which insulin is implicated. The peptides are used to identify agents  
CC that are potentially useful for treating insulin-associated diseases,  
CC particularly diabetes and obesity but also polycystic ovarian syndrome  
CC and syndrome X.

XX Sequence 80 AA;

XX Query Match 75.0%; Score 159; DB 21; Length 80;  
XX Best Local Similarity 69.8%; Pred. No. 2.3e-16;  
XX Matches 30; Conservative 6; Mismatches 7; Indels 0; Gaps 0;

OY 1 PMSISENSLVAMDFSGOKSRVIENTPEALSVAVEGLAWRK 43  
DB 13 PLSVSDNTLVAMDFSGHAGRVINDNPREALSAAEQAQRKK 55

RESULT 14  
AAB18962  
ID AAB18962 standard; peptide: 80 AA.

XX AAB18962;  
XX  
XX 08-FEB-2001 (first entry)  
XX  
XX Peptide derived from the PIR or PIR-SH2 domain of Grb7 protein.

XX Phosphorylated insulin receptor interacting region; Grb7 family protein;  
KW insulin receptor; tyrosine kinase; insulin; insulin-associated disease;  
KW diabetes; obesity; polycystic ovarian syndrome; syndrome X.

OS Mus muris.

XX WO200055634-A1.

XX 21-SEP-2000.

XX 14-MAR-2000; 2000WO-FR00613.

XX 15-MAR-1999; 99FR-0003159.

XX (CNRS ) CNRS CENT NAT RECH SCI.

PI Burnol A, Perdereau D, Kasus-Jacobi A, Bereziat V, Girard J;

DR WPI; 2000-587566/55.

XX Fragments of Grb family proteins to identify compounds are useful in  
PT treating insulin-associated diseases, particularly diabetes and obesity  
PT

PS Claim 2; Page 37; 46pp; French.

XX BI8937-64 represent the PIR (phosphorylated insulin receptor interacting  
CC region) or PIR-SH2 (Src homology 2) domains of a Grb7 family protein.  
CC PIR is the actual binding region but its effect is about 10 times  
CC greater in presence of SH2 (which by itself is inactive). Agents that  
CC affect binding between the peptides and the insulin receptor can  
CC stimulate or inhibit tyrosine kinase activity of the receptor. The  
CC peptides are used for screening molecules for ability to treat diseases  
CC in which insulin is implicated. The peptides are used to identify agents  
CC that are potentially useful for treating insulin-associated diseases,  
CC particularly diabetes and obesity but also polycystic ovarian syndrome  
CC and syndrome X.

XX Sequence 80 AA;

XX Query Match 75.0%; Score 159; DB 21; Length 80;  
XX Best Local Similarity 69.8%; Pred. No. 2.3e-16;  
XX Matches 30; Conservative 6; Mismatches 7; Indels 0; Gaps 0;

OY 1 PMSISENSLVAMDFSGOKSRVIENTPEALSVAVEGLAWRK 43  
DB 13 PLSVSDNTLVAMDFSGHAGRVINDNPREALSAAEQAQRKK 55

RESULT 15  
ABP08708  
ID ABP08708 standard; Protein: 67 AA.

XX

AAC76197.

```

Query Match      23.1%; Score 49; DB 23; Length 67;
Best Local Similarity 33.3%; Pred No. 11;
Matches 12; Conservative 3; Mismatches 15; Indels 6; Gaps 1;

QY  ..11 VAMDFGQKSRVNIEN-----PTALSAVVEGLAW 40
      : |||||::: | | | | | | | | | |

```

XX 03-SEP-2001 (first entry)  
 XX Human colon cancer antigen protein SEQ ID NO:6961.  
 DE Human colon cancer antigen; diagnosis; detection;  
 XX Human; colon cancer; colon cancer antigen; diagnosis; detection;  
 KW Colorectal carcinoma.  
 XX Homo sapiens.  
 XX WO200122920-A2.  
 XX 05-APR-2001.  
 XX 28-SEP-2000; 2000WO-US26524.  
 XX 29-SEP-1999; 99US-0157137.  
 XX 03-NOV-1999; 99US-0163280.  
 XX (HUMA-) HUMAN GENOME-SCI INC.  
 XX Ruben SM, Barash SC, Birse CE, Rosen CA;  
 XX WPI; 2001-235357/24.  
 XX N-PSDB; AAH35602.  
 XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,  
 PT useful for preventing, diagnosing and/or treating colorectal cancers -  
 XX Claim 11; Page 8390; 9803pp; English.  
 XX AAH32943 to AAH37195 and AAH37196 to AAH37788 represent human colon  
 CC cancer-associated nucleic acid molecules (N) and proteins (P), where  
 CC the proteins are collectively known as colon cancer antigens. The colon  
 CC cancer antigens have cytostatic activity and can be used in gene  
 CC therapy and vaccine production. N and P may be used in the prevention,  
 CC diagnosis and treatment of diseases associated with inappropriate P  
 CC expression. For example, N and P may be used to treat disorders  
 CC associated with decreased expression by rectifying mutations or deletions  
 CC in a patient's genome that affect the activity of P by expressing  
 CC inactive proteins or to supplement the patients own production of P.  
 CC Additionally, N may be used to produce the colon cancer-associated P,  
 CC by inserting the nucleic acids into a host cell and culturing the cell  
 CC to express the proteins. N and P can be used in the prevention, diagnosis  
 CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204  
 CC and AAH37789 represent sequences used in the exemplification of the  
 CC present invention. 662 and page 7053 of the sequence listing were  
 CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were  
 CC missing at time of publication, meaning no sequences are present for  
 CC SEQ ID NO:1027 to 1052, 7921 and 7922.  
 XX Sequence 73 AA;  
 Query Match 22.4%; Score 47.5; DB 22; Length 73;  
 Best Local Similarity 30.4%; Pred. No. 20;  
 Matches 14; Conservative 8; Mismatches 17; Indels 7; Gaps 2;  
 OY 4 SISENSLIVAMDFSGKSRVIE-----NPTAL--SVAVEGLAWRK 42  
 DB 11 TISENLFATTGYPGKMASQFIHHLGHPQILMGSVAVGSLSWHR 56  
 RESULT 18  
 ID AEP02324 standard; Protein; 57 AA.  
 XX AEP02324;  
 XX 24-JUN-2002 (first entry)  
 XX Human ORFX protein sequence SEQ ID NO:4630.  
 DE Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;  
 KW

KW hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;  
 KW degenerative disorder; osteoarthritis; neurodegenerative disorder;  
 KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;  
 KW hypertension; hypothyroidism; cholesterol ester storage disease;  
 KW immune deficiency; immune disorder; infectious disease;  
 KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;  
 XX myasthenia gravis.  
 XX Homo sapiens.  
 XX WO2001292523-A2.  
 XX 06-DEC-2001.  
 XX 29-MAY-2001; 2001WO-US10836.  
 XX 30-MAY-2000; 2000US-206132P.  
 XX 29-AUG-2000; 2000US-228716P.  
 XX (CURA-) CURAGEN CORP.  
 XX Shinkets RA, Leach MD;  
 XX WPI; 2002-106308/14.  
 XX N-PSDB; ABN18076.  
 XX Novel human polypeptides and polynucleotides useful for diagnosing,  
 PT preventing and treating cardiovascular disease, neurodegenerative,  
 PT hyperproliferative disorders and autoimmune disorders  
 XX Disclosure; SEQ ID 4630; 1037pp; English.  
 XX The present invention describes substantially purified human proteins  
 CC (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1  
 CC in the specification). ABN15762 to ABN27352 encode the human ORFX  
 CC proteins given in AEP00010 to AEP11500. ORFX proteins are useful for  
 CC treating or preventing a pathology associated with an ORFX-associated  
 CC syndrome in humans, and in the manufacture of a medicament for treating a  
 CC syndrome associated with ORFX-associated disorder. ORFX polynucleotide  
 CC sequences can be used in gene therapy. ORFX sequences can be used in the  
 CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver,  
 CC psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,  
 CC osteoarthritis, neurodegenerative disorders, diabetes mellitus, systemic  
 CC transplantation, cardiovascular diseases, hypothyroidism, cholesterol ester  
 CC storage disease, various immune deficiencies and disorders, rheumatoid  
 CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid  
 CC arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host  
 CC disease and autoimmune inflammatory eye disease. ORFX proteins are also  
 CC useful for treating burns, incisions, ulcers, for treating osteoporosis,  
 CC bone degenerative disorders, or periodontal disease, and for gut  
 CC protection or regeneration and treatment of lung or liver fibrosis,  
 CC reperfusion injury in various tissues and conditions resulting from  
 CC systemic cytokine damage.  
 CC N.B. The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX Sequence 57 AA;  
 Query Match 21.5%; Score 45.5; DB 23; Length 57;  
 Best Local Similarity 36.4%; Pred. No. 29;  
 Matches 12; Conservative 7; Mismatches 5; Indels 9; Gaps 2;  
 OY 15 FSGKSRVIEPT-----EALSVAVEGLAWRK 41  
 DB 17 WSGO---VLENAVRLRREPLNLSLQNGKSWR 46  
 RESULT 19  
 ID ABG47266 standard; Peptide; 84 AA.  
 XX

AC ABC47266;  
 XX  
 DT 19-AUG-2002 (first entry)  
 XX  
 DE Human peptide encoded by genome-derived single exon probe SEQ ID 36931.  
 XX  
 KW Human; single exon probe; asthma; lung cancer; COPD; ILD;  
 KW chronic obstructive pulmonary disease; interstitial lung disease;  
 KW familial idiopathic pulmonary fibrosis; neurofibromatosis;  
 KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;  
 KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemorrhoidosis;  
 KW pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome;  
 KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;  
 KW primary ciliary dyskinesia; pulmonary hypertension;  
 KW hyaline membrane disease.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200186003-A2.  
 XX  
 PD 15-NOV-2001.  
 XX  
 PF 30-JAN-2001; 2001WO-US00665.  
 XX  
 PR 04-FEB-2000; 2000US-180312P.  
 PR 26-MAY-2000; 2000US-207456P.  
 PR 30-JUN-2000; 2000US-060840B.  
 PR 03-AUG-2000; 2000US-0632366.  
 PR 21-SEP-2000; 2000US-234687P.  
 PR 27-SEP-2000; 2000US-236359P.  
 PR 04-OCT-2000; 2000GB-0024263.  
 XX  
 FA (MOLE-) MOLECULAR DYNAMICS INC.  
 XX  
 PI Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX  
 DR WPI; 2002-114183/15.  
 XX  
 PT Spatially-addressable set of single exon nucleic acid probes, used to  
 XX measure gene expression in human lung samples -  
 XX  
 PS Claim 27; SEQ ID NO 36931; 634pp; English.  
 XX  
 CC The invention relates to a spatially-addressable set of single exon  
 CC nucleic acid probes for measuring gene expression in a sample derived  
 CC from human lung comprising single exon nucleic acid probes having one of  
 CC 12614 nucleic acid sequences mentioned in the specification, or their  
 CC complements or the 12387 open reading frames derived from the 12614  
 CC probes. Also included are a microarray comprising the novel set of  
 CC probes; the novel set of probes which hybridise at high stringency to a  
 CC nucleic acid expressed in the human lung; measuring gene expression in a  
 CC sample derived from human lung, comprising (a) contacting the array with  
 CC a collection of detectably labeled nucleic acids derived from human lung  
 CC mRNA, and (b) measuring the label detectably bound to each probe of  
 CC the array; identifying exons in a eukaryotic genome, comprising  
 CC (a) algorithmically predicting at least one exon from genomic sequences  
 CC of the eukaryote; and (b) detecting specific hybridisation of detectably  
 CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,  
 CC in having a fragment identical to the predicted exon, the probe is included  
 CC in the above mentioned microarray; assigning exons to a single gene,  
 CC comprising (a) identifying exons from genomic sequence by the method  
 CC above and (b) measuring the expression of each of the exons in several  
 CC tissues and/or cell types using hybridisation to a single exon  
 CC microarrays having a probe with the exon, where a common pattern of  
 CC expression of the exons in the tissues and/or cell types indicates that  
 CC the exons should be assigned to a single gene; a peptide comprising one  
 CC of 12011 sequences, mentioned in the specification, or encoded by the  
 CC probes/open reading frames (ORF). The probes are used for gene  
 CC expression analysis, and for identifying exons in a gene, particularly  
 CC such as human lung derived mRNA and for the study of lung diseases  
 CC including asthma, lung cancer, chronic obstructive pulmonary disease  
 CC (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary  
 CC fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease,

CC Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary  
 CC haemosiderosis, pulmonary histiocytosis, lymphangioleiomyomatosis,  
 CC pulmonary alveolar proteinosis, Karagener syndrome, fibrocystic  
 CC and hyaline membrane disease, primary ciliary dyskinesia, pulmonary hypertension  
 CC encoded by a single exon probe of the invention.  
 CC Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 84 AA;  
 Query Match 21.2%; Score 45; DB 23; Length 84;  
 Best Local Similarity 26.4%; Pred. No. 59;  
 Matches 14; Conservative 6; Mismatches 7; Indels 26; Gaps 2;  
 QY 17 GOKSRVIENP-----TEALSVAV-----EEGLARRKK 43  
 DB 11 GOKARLSRPLRGVSKHCLTFYHMYGGTGLLSVYLKKEDESESLWRRR 63  
 RESULT 20  
 AAG03340  
 ID AAG03340 standard; Protein; 72 AA.  
 XX  
 AC AAG03340;  
 XX  
 DT 06-OCT-2000 (first entry)  
 XX  
 DE Human secreted protein, SEQ ID NO: 7421.  
 KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
 KW gene therapy; chromosome mapping.  
 XX  
 OS Homo sapiens.  
 XX  
 PN EP1033401-A2.  
 XX  
 PD 06-SEP-2000.  
 XX  
 PF 21-FEB-2000; 2000EP-0200610.  
 PR 26-FEB-1999; 99US-0122487.  
 XX  
 PA (GEST ) GENSET.  
 XX  
 PI Dumas Milne Edwards J, Duclert A, Giordano J;  
 DR WPI; 2000-500381/45.  
 DR N-PSDB; AAC03346.  
 XX  
 PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
 PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for  
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -  
 XX  
 PS Claim 13; SEQ ID 7421; 71pp + CD-ROM; English.  
 CC  
 CC The present sequence is a polypeptide encoded by one of a large number  
 CC of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs  
 CC were prepared from total human RNAs or polyA+ RNAs derived from 30  
 CC different tissues. EST sequences usually correspond mainly to the 3'  
 CC untranslated region (UTR) of the mRNA because they are often obtained  
 CC from oligo-dT primed cDNA libraries. Such ESTs are not well suited for  
 CC isolating cDNA sequences derived from the 5' ends of mRNAs and even in  
 CC those cases where longer cDNA sequences have been obtained, the full 5'  
 CC UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5'  
 CC ends and can therefore be used to obtain full length cDNAs and genomic  
 CC DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and  
 CC chromosome mapping procedures. They are used to obtain upstream  
 CC regulatory sequences and to design expression and secretion vectors.  
 XX  
 SQ Sequence 72 AA;

Query Match 21.0%; Score 44.5; DB 21; Length 72;  
 Best Local Similarity 50.0%; Pred. No. 57;  
 Matches 10; Conservative 4; Mismatches 5; Indels 1; Gaps 1;

QY 21 RVIENTPALSVAVEGLAW 40  
 || :|||: ||| :|: |  
 Db 38 RVCPTSCSVA-QAGVQW 56

RESULT 21  
 AAU87164  
 ID AAU87164 standard; Protein; 74 AA.

XX AC AAU87164;

XX DT 05-JUN-2002 (first entry)

XX DE Novel central nervous system protein #74.

XX KW Central nervous system; CNS; autoimmune disease; rheumatoid arthritis;  
 KW hyperproliferative disorder; neoplasm; cardiovascular disorder;  
 KW cardiac arrest; cerebrovascular disorder; ischaemia; angiogenesis;  
 KW nervous system disorder; Alzheimer's disease; AIDS; ocular disorder;  
 KW acquired immunodeficiency virus; dysphagia; gastrointestinal disorder;  
 KW adenocarcinoma; reproductive system disorder; testicular feminisation;  
 KW endocrine disorder; diabetes; cancer; leukaemia; neovascularisation;  
 KW respiratory disorder; renal disorder; kidney failure; blood disorder;  
 KW myocardial infarction; wound healing; cell proliferation; skin aging;  
 KW food additive; food preservative; gene therapy.

XX OS Homo sapiens.

XX PN WO200155318-A2.

XX PD 02-AUG-2001.

XX PF 17-JAN-2001; 2001WO-US01332.

XX 31-JAN-2000; 2000US-0179065.  
 PR 04-FEB-2000; 2000US-0180628.  
 PR 24-FEB-2000; 2000US-0184664.  
 PR 02-MAR-2000; 2000US-0186350.  
 PR 16-MAR-2000; 2000US-0189874.  
 PR 17-MAR-2000; 2000US-0190076.  
 PR 18-APR-2000; 2000US-0198123.  
 PR 19-MAY-2000; 2000US-0205515.  
 PR 07-JUN-2000; 2000US-0209467.  
 PR 28-JUN-2000; 2000US-0214886.  
 PR 30-JUN-2000; 2000US-0215135.  
 PR 07-JUL-2000; 2000US-0216647.  
 PR 07-JUL-2000; 2000US-0216880.  
 PR 11-JUL-2000; 2000US-0217487.  
 PR 11-JUL-2000; 2000US-0217496.  
 PR 14-JUL-2000; 2000US-0218290.  
 PR 26-JUL-2000; 2000US-0220963.  
 PR 26-JUL-2000; 2000US-0220964.  
 PR 14-AUG-2000; 2000US-0224518.  
 PR 14-AUG-2000; 2000US-0224519.  
 PR 14-AUG-2000; 2000US-0225213.  
 PR 14-AUG-2000; 2000US-0225214.  
 PR 14-AUG-2000; 2000US-0225266.  
 PR 14-AUG-2000; 2000US-0225267.  
 PR 14-AUG-2000; 2000US-0225268.  
 PR 14-AUG-2000; 2000US-0225270.  
 PR 14-AUG-2000; 2000US-0225277.  
 PR 14-AUG-2000; 2000US-0225757.  
 PR 14-AUG-2000; 2000US-0225758.  
 PR 14-AUG-2000; 2000US-0225759.  
 PR 18-AUG-2000; 2000US-0226279.  
 PR 22-AUG-2000; 2000US-0226681.  
 PR 22-AUG-2000; 2000US-0226688.  
 PR 22-AUG-2000; 2000US-0227182.

PR 23-AUG-2000; 2000US-0227009.  
 PR 30-AUG-2000; 2000US-0228924.  
 PR 01-SEP-2000; 2000US-0229287.  
 PR 01-SEP-2000; 2000US-0229343.  
 PR 01-SEP-2000; 2000US-0229344.  
 PR 01-SEP-2000; 2000US-0229345.  
 PR 05-SEP-2000; 2000US-0229509.  
 PR 05-SEP-2000; 2000US-0229513.  
 PR 08-SEP-2000; 2000US-0230437.  
 PR 08-SEP-2000; 2000US-0230438.  
 PR 08-SEP-2000; 2000US-0231242.  
 PR 08-SEP-2000; 2000US-0231243.  
 PR 08-SEP-2000; 2000US-0231244.  
 PR 08-SEP-2000; 2000US-0231413.  
 PR 08-SEP-2000; 2000US-0231414.  
 PR 08-SEP-2000; 2000US-0232080.  
 PR 08-SEP-2000; 2000US-0232081.  
 PR 12-SEP-2000; 2000US-0231968.  
 PR 14-SEP-2000; 2000US-0232397.  
 PR 14-SEP-2000; 2000US-0232398.  
 PR 14-SEP-2000; 2000US-0232399.  
 PR 14-SEP-2000; 2000US-0232400.  
 PR 14-SEP-2000; 2000US-0232401.  
 PR 14-SEP-2000; 2000US-0233063.  
 PR 14-SEP-2000; 2000US-0233064.  
 PR 14-SEP-2000; 2000US-0233065.  
 PR 21-SEP-2000; 2000US-0234323.  
 PR 21-SEP-2000; 2000US-0234327.  
 PR 23-SEP-2000; 2000US-0234397.  
 PR 23-SEP-2000; 2000US-0234598.  
 PR 26-SEP-2000; 2000US-0235484.  
 PR 27-SEP-2000; 2000US-0235834.  
 PR 29-SEP-2000; 2000US-0236327.  
 PR 29-SEP-2000; 2000US-0236367.  
 PR 29-SEP-2000; 2000US-0236368.  
 PR 29-SEP-2000; 2000US-0236369.  
 PR 29-SEP-2000; 2000US-0236370.  
 PR 02-OCT-2000; 2000US-0236802.  
 PR 02-OCT-2000; 2000US-0237037.  
 PR 02-OCT-2000; 2000US-0237038.  
 PR 02-OCT-2000; 2000US-0237039.  
 PR 02-OCT-2000; 2000US-0237040.  
 PR 13-OCT-2000; 2000US-0238935.  
 PR 13-OCT-2000; 2000US-0238937.  
 PR 20-OCT-2000; 2000US-0240960.  
 PR 20-OCT-2000; 2000US-0241221.  
 PR 20-OCT-2000; 2000US-0241785.  
 PR 20-OCT-2000; 2000US-0241786.  
 PR 20-OCT-2000; 2000US-0241787.  
 PR 20-OCT-2000; 2000US-0241808.  
 PR 20-OCT-2000; 2000US-0241826.  
 PR 01-NOV-2000; 2000US-0244617.  
 PR 08-NOV-2000; 2000US-0246474.  
 PR 08-NOV-2000; 2000US-0246475.  
 PR 08-NOV-2000; 2000US-0246476.  
 PR 08-NOV-2000; 2000US-0246477.  
 PR 08-NOV-2000; 2000US-0246478.  
 PR 08-NOV-2000; 2000US-0246523.  
 PR 08-NOV-2000; 2000US-0246524.  
 PR 08-NOV-2000; 2000US-0246525.  
 PR 08-NOV-2000; 2000US-0246526.  
 PR 08-NOV-2000; 2000US-0246537.  
 PR 08-NOV-2000; 2000US-0246538.  
 PR 08-NOV-2000; 2000US-0246539.  
 PR 08-NOV-2000; 2000US-0246609.  
 PR 08-NOV-2000; 2000US-0246610.  
 PR 08-NOV-2000; 2000US-0246611.  
 PR 08-NOV-2000; 2000US-0246613.  
 PR 17-NOV-2000; 2000US-0249207.  
 PR 17-NOV-2000; 2000US-0249208.  
 PR 17-NOV-2000; 2000US-0249209.

4 SISENSLVAMDFSGQKSRVIEN--PTEALSVA-----VEEGLAWKK 43  
 ||:|| | : ||: ||: | | : :  
 : ||| : ||| :

20 SITE: ENGLIPKDYRSLKTOYLSQSGPEHLFTSNLRXAGLLTXOAPGDNXTAWRYK 74

RESULT 22  
AAU087480  
ID AAU087480 standard; Protein; 74 AA.  
XX  
XX AAU087480;  
XX  
DT 05-JUN-2002 (first entry)  
XX  
DE Novel central nervous system protein #390.  
XX  
KW Central nervous system; CNS; autoimmune disease; rheumatoid arthritis;  
KW hyperproliferative disorder; neoplasm; cardiovascular disorder;  
KW cardiac arrest; cerebrovascular disorder; ischaemia; angiogenesis;  
KW nervous system disorder; Alzheimer's disease; AIDS; ocular disorder;  
KW acquired immunodeficiency virus; dysphagia; gastrointestinal disorder;  
KW adenocarcinoma; reproductive virus; system disorder; testicular feminisation;  
KW endocrine disorder; diabetes; cancer; leukaemia; neovascularisation;  
KW respiratory disorder; renal disorder; kidney failure; blood disorder;  
KW myocardial infarction; wound healing; cell proliferation; skin aging;  
KW food additive; food preservative; gene therapy.  
XX  
OS Homo sapiens.  
XX  
PN WO200155318-A2.  
XX  
PD 02-AUG-2001.  
XX  
PF 17-JAN-2001; 2001WO-US01332.  
XX  
PR 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 15-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205513.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217496.  
PR 11-JUL-2000; 2000US-0217497.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 22-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226686.  
PR 23-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0228927.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.

PR 05-SEP-2000; 2000US-0229513.  
 PR 06-SEP-2000; 2000US-0230437.  
 PR 06-SEP-2000; 2000US-0230438.  
 PR 08-SEP-2000; 2000US-02311242.  
 PR 08-SEP-2000; 2000US-02311243.  
 PR 08-SEP-2000; 2000US-02311244.  
 PR 08-SEP-2000; 2000US-02311413.  
 PR 08-SEP-2000; 2000US-02311414.  
 PR 08-SEP-2000; 2000US-0232080.  
 PR 08-SEP-2000; 2000US-0232081.  
 PR 12-SEP-2000; 2000US-0231968.  
 PR 14-SEP-2000; 2000US-0232397.  
 PR 14-SEP-2000; 2000US-0232398.  
 PR 14-SEP-2000; 2000US-0232399.  
 PR 14-SEP-2000; 2000US-0232400.  
 PR 14-SEP-2000; 2000US-0233063.  
 PR 14-SEP-2000; 2000US-0233064.  
 PR 14-SEP-2000; 2000US-0233065.  
 PR 21-SEP-2000; 2000US-0234223.  
 PR 21-SEP-2000; 2000US-0234274.  
 PR 25-SEP-2000; 2000US-0234997.  
 PR 25-SEP-2000; 2000US-0234998.  
 PR 26-SEP-2000; 2000US-0235484.  
 PR 27-SEP-2000; 2000US-0235834.  
 PR 27-SEP-2000; 2000US-0235836.  
 PR 29-SEP-2000; 2000US-0236327.  
 PR 29-SEP-2000; 2000US-0236367.  
 PR 29-SEP-2000; 2000US-0236368.  
 PR 29-SEP-2000; 2000US-0236369.  
 PR 29-SEP-2000; 2000US-0236370.  
 PR 02-OCT-2000; 2000US-0236802.  
 PR 02-OCT-2000; 2000US-0237037.  
 PR 02-OCT-2000; 2000US-0237038.  
 PR 02-OCT-2000; 2000US-0237039.  
 PR 02-OCT-2000; 2000US-0237040.  
 PR 13-OCT-2000; 2000US-0239935.  
 PR 13-OCT-2000; 2000US-0239937.  
 PR 20-OCT-2000; 2000US-0240960.  
 PR 20-OCT-2000; 2000US-0241221.  
 PR 20-OCT-2000; 2000US-0241785.  
 PR 20-OCT-2000; 2000US-0241786.  
 PR 20-OCT-2000; 2000US-0241787.  
 PR 20-OCT-2000; 2000US-0241808.  
 PR 20-OCT-2000; 2000US-0241809.  
 PR 20-OCT-2000; 2000US-0241826.  
 PR 01-NOV-2000; 2000US-0244617.  
 PR 08-NOV-2000; 2000US-0246474.  
 PR 08-NOV-2000; 2000US-0246475.  
 PR 08-NOV-2000; 2000US-0246476.  
 PR 08-NOV-2000; 2000US-0246477.  
 PR 08-NOV-2000; 2000US-0246478.  
 PR 08-NOV-2000; 2000US-0246523.  
 PR 08-NOV-2000; 2000US-0246524.  
 PR 08-NOV-2000; 2000US-0246525.  
 PR 08-NOV-2000; 2000US-0246526.  
 PR 08-NOV-2000; 2000US-0246527.  
 PR 08-NOV-2000; 2000US-0246528.  
 PR 08-NOV-2000; 2000US-0246532.  
 PR 08-NOV-2000; 2000US-0246609.  
 PR 08-NOV-2000; 2000US-0246610.  
 PR 08-NOV-2000; 2000US-0246611.  
 PR 08-NOV-2000; 2000US-0246613.  
 PR 17-NOV-2000; 2000US-0249207.  
 PR 17-NOV-2000; 2000US-0249208.  
 PR 17-NOV-2000; 2000US-0249210.  
 PR 17-NOV-2000; 2000US-0249211.  
 PR 17-NOV-2000; 2000US-0249212.  
 PR 17-NOV-2000; 2000US-0249213.  
 PR 17-NOV-2000; 2000US-0249214.  
 PR 17-NOV-2000; 2000US-0249215.  
 PR 17-NOV-2000; 2000US-0249216.

PR 17-NOV-2000; 2000US-0249217.  
 PR 17-NOV-2000; 2000US-0249218.  
 PR 17-NOV-2000; 2000US-0249244.  
 PR 17-NOV-2000; 2000US-0249245.  
 PR 17-NOV-2000; 2000US-0249264.  
 PR 17-NOV-2000; 2000US-0249265.  
 PR 17-NOV-2000; 2000US-0249267.  
 PR 17-NOV-2000; 2000US-0249269.  
 PR 01-DEC-2000; 2000US-0251000.  
 PR 01-DEC-2000; 2000US-0250391.  
 PR 05-DEC-2000; 2000US-0251030.  
 PR 05-DEC-2000; 2000US-0251298.  
 PR 05-DEC-2000; 2000US-0251719.  
 PR 06-DEC-2000; 2000US-0251719.  
 PR 08-DEC-2000; 2000US-0251856.  
 PR 08-DEC-2000; 2000US-0251868.  
 PR 08-DEC-2000; 2000US-0251869.  
 PR 08-DEC-2000; 2000US-0251869.  
 PR 11-DEC-2000; 2000US-0251890.  
 PR 05-JAN-2001; 2001US-0259678.  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 XX Rosen CA, Barash SC, Ruben SM;  
 PI WPI: 2001-581633/65.  
 DR N-PSDB; ABK43810.  
 XX New isolated nucleic acid encoding a protein for diagnosing,  
 PT preventing, treating or ameliorating medical conditions and used as  
 PT food additives or preservatives -  
 PS Claim 9; SEQ ID NO 998; 837pp; English.  
 XX The invention describes an isolated nucleic acid molecule (i) encoding a  
 CC novel central nervous system protein. (ii) and polypeptides (iii) encoded  
 CC by (i), are used to treat a medical conditions and in diagnosis of a  
 CC pathological condition. Disorders which are diagnosed or treated include  
 CC autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative  
 CC disorders e.g. neoplasms of the breast or liver, cardiovascular disorders  
 CC e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,  
 CC anglogenesis, nervous system disorders e.g. Alzheimer's disease and  
 CC amyotrophic lateral sclerosis, infections caused by bacteria, viruses  
 CC e.g. Acquired immunodeficiency virus (AIDS) and fungi, ocular disorders  
 CC e.g. corneal infection, gastrointestinal disorders e.g. dysphagia,  
 CC adenocarcinomas and irritable bowel syndrome, endocrine disorders e.g. diabetes  
 CC and pituitary dwarfism, cancers and disorders at the cellular level e.g.  
 CC leukaemia, disorders involving neovascularisation e.g. malignancies,  
 CC acute kidney failure and blood related disorders e.g. myocardial  
 CC infarction. The polypeptides can also be used to aid wound healing and  
 CC epithelial cell proliferation, to prevent skin aging due to sunburn, to  
 CC maintain organs before transplantation, for supporting cell culture of  
 CC primary tissues, to regenerate tissues and in chemotaxis. The  
 CC polypeptides can also be used as a food additive or preservative to  
 CC increase or decrease storage capabilities, fat content, lipid, protein,  
 Query Match 21.0%; Score 44.5; DB 22; Length 74;  
 Best Local Similarity 23.6%; Pred. No. 59;  
 Matches 13; Conservative 11; Mismatches 16; Indels 15; Gaps 2;  
 OY 4 SISENSLVAMDFSGQSRVIEV-PTAALSV-VEGLAWRK 43  
 DB 20 SITENGLIPKDYRLATQYLSYGPHELLTFLSLRXAGLLTQXAPGDNTAWRVK 74  
 RESULT 23  
 AAU41349  
 ID AAU41349 standard; Protein; 79 AA.  
 XX



```

AC AAU41349;
XX
DN 13-FEB-2002 (first entry)
XX
DE Propionibacterium acnes immunogenic protein #2245.
XX
KW SAPHO syndrome; synovitis; acne; pustulosis; hyperteositis; osteomyelitis;
KW uvelitis; endophthalmitis; bone; joint; central nervous system; ELISA;
KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
KW dermatological; osteopathic; neuroprotectant.
XX
OS Propionibacterium acnes.
XX
PN WQ2000181581-A2.
XX
PD 01-NOV-2001.
XX
PF 20-APR-2001; 2001WO-US12865.
XX
PR 21-APR-2000; 2000US-199047P.
XX
PR 02-JUN-2000; 2000US-208841P.
XX
PR 07-JUL-2000; 2000US-216747P.
XX
PA (CORI-) CORIXA CORP.
XX
PI Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
PI L'maisonneuve J, Zhang Y, Jen S, Carter D;
XX
DR WPI: 2001-616774/71.
XX
DR N-PSDB; AAS59515.
XX
PT Propionibacterium acnes polypeptides and nucleic acids useful for
PT vaccinating against and diagnosing infections, especially useful for
PT treating acne vulgaris -
XX
PS Example 1; SEQ ID NO 2544; 1069pp; English.
XX
CC Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
CC polypeptides. The proteins and their associated DNA sequences are used in
CC the treatment, prevention and diagnosis of medical conditions caused by
CC P. acnes. The disorders include SAPHO syndrome, synovitis, acne,
CC pustulosis, hypertostis and osteomyelitis), uvelitis, arthritis, acne,
CC P. acnes is also involved in infections of bone joints and the central
CC nervous system, however it is particularly involved in the inflammatory
CC lesions associated with acne vulgaris. A method for detecting the
CC presence or absence of P. acnes in a patient comprises contacting a
CC sample with a binding agent that binds to the proteins of the invention
CC and determining the amount of bound protein in the sample. The
CC polypeptides may be used as antigens in the production of antibodies
CC specific for P. acnes proteins. These antibodies can be used to
CC downregulate expression and activity of P. acnes polypeptides and
CC therefore treat P. acnes infections. The antibodies may also be used as
CC diagnostic agents for determining P. acnes presence, for example, by
CC enzyme linked immunosorbent assay (ELISA).
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 79 AA;
Query Match
Best Local Similarity 21.0%; Score 44.5; DB 22; Length 79;
Matches 11; Conservative 10; Mismatches 16; Indels 3; Gaps 1;
QY 4 STSENLSLVAMPDSGQSRYVTENPTALSVAVEGLAWKK 43
| : | : | : | : | : | : | : | : | : | :
Db 28 SLVNSVPVALSREGPSNRV---PTRSLCATRRHGVCSRER 64
RESULT 24
AAU62249
ID ID AAU62249 standard; Protein: 90 AA
```

[illegible]

AC AAW94366;  
XX 21-NOV-2001 (first entry)  
XX Human reproductive system related antigen SEQ ID NO: 3024.  
XX Human; reproductive system related antigen; reproductive system disorder;  
XX cancer; gene therapy.  
XX Homo sapiens.  
XX WO200155320-A2.  
XX 02-AUG-2001.  
XX 17-JAN-2001; 2001WO-US01339.  
XX 31-JAN-2000; 2000US-0179055.  
XX 04-FEB-2000; 2000US-0180628.  
XX 24-FEB-2000; 2000US-0184664.  
XX 02-MAR-2000; 2000US-0186350.  
XX 16-MAR-2000; 2000US-0189874.  
XX 17-MAR-2000; 2000US-0190076.  
XX 18-APR-2000; 2000US-0198123.  
XX 19-MAY-2000; 2000US-0205515.  
XX 07-JUN-2000; 2000US-0209467.  
XX 28-JUN-2000; 2000US-0214886.  
XX 30-JUN-2000; 2000US-0215135.  
XX 07-JUL-2000; 2000US-0216647.  
XX 07-JUL-2000; 2000US-0216880.  
XX 11-JUL-2000; 2000US-0217487.  
XX 14-JUL-2000; 2000US-0218290.  
XX 26-JUL-2000; 2000US-0220963.  
XX 26-JUL-2000; 2000US-0220964.  
XX 14-AUG-2000; 2000US-0224518.  
XX 14-AUG-2000; 2000US-0224519.  
XX 14-AUG-2000; 2000US-0225213.  
XX 14-AUG-2000; 2000US-0225214.  
XX 14-AUG-2000; 2000US-0225266.  
XX 14-AUG-2000; 2000US-0225267.  
XX 14-AUG-2000; 2000US-0225268.  
XX 14-AUG-2000; 2000US-0225447.  
XX 14-AUG-2000; 2000US-0225757.  
XX 14-AUG-2000; 2000US-0225758.  
XX 18-AUG-2000; 2000US-0226379.  
XX 22-AUG-2000; 2000US-0226681.  
XX 22-AUG-2000; 2000US-0226686.  
XX 22-AUG-2000; 2000US-0227182.  
XX 23-AUG-2000; 2000US-0227009.  
XX 30-AUG-2000; 2000US-0228924.  
XX 01-SEP-2000; 2000US-0229287.  
XX 01-SEP-2000; 2000US-0229343.  
XX 01-SEP-2000; 2000US-0229344.  
XX 05-SEP-2000; 2000US-0229345.  
XX 05-SEP-2000; 2000US-0229509.  
XX 05-SEP-2000; 2000US-0229513.  
XX 06-SEP-2000; 2000US-0230437.  
XX 06-SEP-2000; 2000US-0230438.  
XX 08-SEP-2000; 2000US-0231242.  
XX 08-SEP-2000; 2000US-0231243.  
XX 08-SEP-2000; 2000US-0231244.  
XX 08-SEP-2000; 2000US-0231413.  
XX 08-SEP-2000; 2000US-0231414.  
XX 08-SEP-2000; 2000US-0232080.  
XX 08-SEP-2000; 2000US-0232081.  
XX 12-SEP-2000; 2000US-0231968.  
XX 14-SEP-2000; 2000US-0232397.  
XX 14-SEP-2000; 2000US-0232398.  
XX 14-SEP-2000; 2000US-0232399.  
XX 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 21-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 23-SEP-2000; 2000US-0234997.  
PR 23-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.

Fri Mar 28 12:13:32 2003

us-09-936-697-5.max.rag

Page 22

```
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251869.
PR 11-DEC-2000; 2000US-0251890.
PR 05-JAN-2001; 2001US-0254097.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX PI Rosen CA, Barash SC, Ruben SM;
XX DR WPI; 2001-465570/50.
XX DR N-PSDB; AAL00336.
XX
XX Isolated nucleic acid molecule encoding a reproductive system antigen
XX is used in preventing, treating or ameliorating a medical condition -
XX Claim 11; SEQ ID NO 3024; 1297pp + Sequence Listing; English.
XX
XX The present invention provides the protein and coding sequences of a
XX number of human reproductive system related antigens. These can be used
XX in the prevention and treatment of reproductive system disorders,
XX including cancer. The present sequence is a protein of the invention.
XX
XX Sequence 47 AA;
SQ
Query Match
Best Local Similarity 20.8%; Score 44; DB 22; Length 47;
Matches 8; Conservative 5; Mismatches 5; Indels 0; Gaps 0;
OY 20 SRVIENPTAALSVAVEEG 37
Db |||: || :||: |
7 SRVLKGPTNIVSLVNSG 24
Search completed: March 28, 2003, 09:05:30
Job time : 58.8583 secs
```